UNIVERSITY of CALIFORNIA
Santa Barbara

Fault-tolerant superconducting qubits

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Physics

by

Julian S. Kelly

Committee in charge:

Professor John Martinis, Chair
Professor David Weld
Professor Wim van Dam

March 2015

The dissertation of Julian S. Kelly is approved:

Professor David Weld

Professor Wim van Dam

Professor John Martinis, Chair

March 2015

To my parents, Peter and Kathryn.

Acknowledgements

John, first and foremost, thank you for giving me the opportunity to work in the lab, and for teaching me how to perform amazing experiments. When you gave me my first project as an undergraduate, you gave me my first experience to really dig into research; I've been hooked ever since. You've been a great adviser who always pushes for research to be better, clearer, and more polished, yet to never hide the "dirt physics" behind the scene. Your enthusiastic attitude towards physics inspires me and many others. Lastly, thank you for teaching me the "right ways" to take, analyze, and present data.

Rami, working with you has been an incredibly educational and productive experience. We may have gotten off to a little bit of a rocky start, but we really hit our stride and accomplished more than I ever could have expected in the time of my PhD. You haven been a fantastic mentor, co-worker, co-author, and friend. Thanks for your advice in the lab and elsewhere, teaching me to write beautiful papers, and providing cosmic balance to my unwavering optimism with your unflinching pessimism.

Erik, you have probably had a greater effect on my style as a physicist than you know. I'm greatly indebted to you for spending the time to get me started on the right foot as an undergraduate, and staying up until three in the morning working on the DRAG project. I fondly remember a day where I was building some doohickey for the lab, you came by and noticed that I was doing a pretty crappy job. "Dude, take some fucking pride in your work." I have done my best to internalize those words and your style, from making beautiful devices in the clean room, to clear and aesthetically pleasing talks and papers. You really are a pleasure to work with, dude, and I hope that our paths cross in the not-too-distant future. Maybe an aquarium or something.

As an undergraduate in College of Creative Studies, Matthew was my TA. I soon realized that Matthew was an ideal role model; I wanted to know physics just as well as that guy. And when I later joined the Martinis group, there you were again. Thanks for grading my papers, letting me borrow your qubits, and inspiring me to intuitively understand physics and do great experiments.

Erik and Matthew, thanks for convincing me to play hookie and taking me on that hike in Oregon. This memory really stands out to me, and is a great reminder to not get too caught up in work.

Ted, we probably spend too much time together, but I wouldn't have it any other way. From office-mate, to gym-mate and house-mate, you're a great mate, friend. It has been a pleasure working along side you in lab and decompressing with you afterward, and being able to combine forces on the latest experiments has been incredibly cool. Here's to graduating and filling the world with more quantum device magic.

Austin, it has been really great getting to know you over the last year. We're very fortunate that you gave our lab a solid direction to move toward, and having you down the hall has been invaluable. It has been really fun watching a theorist dive into building electronics for the lab. I can't say that's something I ever expected to see. It has been a real pleasure working with you on this latest publication. Here's hoping to one day seeing a full-fledged superconducting surface code quantum computer.

Dan, oh man we have been through the bad and the good together. The three weeks that we spent in front of the DACs trying to track down missing packets still gives me nightmares, and gateCompiler still makes me smile when I look back on how incomprehensibly useful that thing is. I think its interesting when you say that we have compatible working styles; I am fast and get things working, you are methodical and deliberate. Although we clash sometimes, I think that this is a sign of a healthy working partnership. Lastly, I want to thank you for your contributions to the lab: the GHZ FPGA server, dual block, and fast readout have been invaluable.

Josh, I think it was obvious from the second you entered Broida that you belonged in our lab. Even though you're Canadian. Sorry about that. I had a great time showing you and Ted the ol' a-Si phase qubit process, daily water cooler talk, and coffee Sundays outside of lab. Thanks for your hard work and amazing amplifiers that made all of this possible.

Yu, I'd like to take a moment to thank you for some of the ground work that you did which made all of my research possible. Your contributions to the ADC and dispersive readout led the way for the Xmon results. Bringing up the SiO_2 process was critical to all of the recent complex chips in the lab, so thank you. It has been a pleasure working with you the last few years.

Lookit this guy. Pedram, you bring a really special level of trash talk to the lab. That and occasionally baklava. Seeing you fall in love with LabRAD years ago really brought a tear to my eye. I hope you you and it continue to have a productive relationship for years to come. Maybe one day you and Rami will finally sort out which day of the week is for celebrating what person.

I think our lab really lucked out that you wanted to come here, Evan. You bring an incredible level of competence to the lab in both physics and otherwise. Thanks for all of the help with dual block, pulse caching, the new ADC firmware, and everything else around the lab. It has been great being your neighbor downtown, and watching Barnaby grow up. Thanks for all of the delicious smoked meats and home brews.

Jimmy, it was really great seeing you bring the air bridge project to fruition, despite the numerous challenges. You really have the knack for getting something done once you set your mind to it. I have enjoyed talking about random nonsense with you in the lab, and its great seeing the younger generation of graduate students dig into qubit experiments.

Ben, you and I conquered Albuquerque together. Thanks for helping ensure that we didn't get shanked, and for borrowing a pen from the police cruiser so we could get receipts hand scrawled on a napkin for $4 street vendor hot dogs. You're a champ. I hope that when I announce my first child it will be some fraction as amazing as when you announced yours in Baltimore.

Tony "Q", your hard work in materials laid the foundation for my PhD. Thanks for all those times you dropped what you were doing for 48 hours to crank out an MBE wafer. It has been really fantastic watching your tackle some of our fields hardest problems, and everyone is the better for it.

Charles, I think we can agree that we have a "mutual" respect for each other. Watching you distort the Xmon design into your gmon architecture filled me with feelings of disgust and awe.

But in the end, you all really got that thing working. I enjoyed the countless months toiling away in the DR lab as we got our experiments working.

Peter, king of Java. It has been great the last couple of months pushing out this latest manuscript. Thanks for the countless times you dropped what you were doing to help with the qubit sequencer, and humoring my constant barrage of complaining about L-Edit, LaTeX, python, etc. Here's to you finishing up and starting a family.

I never expected to be asked specific questions about my March Meeting talk at graduate student visit day, but anybody that knows Chris would not be surprised. Your "get your hands dirty" attitude, combined with your persistence and methodical approach to physics will get you far. I'm delighted that you chose to join the group, and I can't wait to see what you accomplish.

Amit, you have been a great office mate and a great friend. Thanks for freeing us from the grips of windows server software, and hooking us up with something that actually works. I've really enjoyed getting to know you the last couple of years, and hearing your interesting stories in lab. It has been a joy working with somebody who is so selfless and willing to help.

Jim, thank you for your constant attention to detail, and consistently ferrying paperwork between Broida and CNSI. Your relentless stream of paper comments has helped maintain the level of polish expected from Martinis group publications. I dream of the day we will finally see the artisanal Wenner rain gutter.

Woof ruff bark. Its probably safe to assume that you will never read this, but I could not have asked for a better dog than you, Qubit. Thanks for the constant companionship and support. May you continue to be the mascot of the quantum computing revolution, and guard many offices in the future.

Dunsworth, between rock climbing, surfing, and owning a dog, you know how to live the Santa Barbara lifestyle to the fullest. Thanks for keeping lab atmosphere down to earth, and I look forward to seeing you rock the rest of your PhD like you do your climbing.

Brooks, thanks for your work making 18 billion diff amps, we definitely put those bad boys to work in the setup on Vince. I wish we got a little more time to build up our downtown-campus biker gang, but I couldn't keep up with you. It's been fun getting you started on qubit experiments; I'm excited to see where you go with them.

Io-Chun, I've enjoyed getting to know you over the last few months, whether that be in lab or running into you downtown. I wish you luck in your future endeavors in China, and maybe one day I will be able to drop by and visit.

Trang, thanks for your constant support and encouraging attitude. Your "fashion tips for physicists" have been and continue to be spot on, and you help make my life just a little more French. You have made this last year a real joy, and I'm sure you will continue to do so for many years to come.

Gavin, you're been a great friend and a great housemate. I have really enjoyed our lengthy discussions on physics, academia, and life. May your pot of coffee never run dry.

Andrew, thank you for teaching the fantastic undergraduate and graduate quantum classes which helped to inspire my interest in the subject. Your challenging classes, challenging advancement questions, and on-point paper comments have encouraged me to strive for the best.

Mom and Dad, thanks for always believing in me and supporting me. Your hard work in pushing me forward and dealing with my difficult teenage years have made me optimistic about everything to come.

Curriculum Vitæ

Julian S. Kelly

Education

(2015) Ph.D., Physics, University of California Santa Barbara

(2010) B.S., Physics, University of California Santa Barbara

(2006) Garfield high school, Seattle, WA

First author publications

"Coherent Josephson qubit suitable for scalable quantum integrated circuits", Rami Barends, Julian Kelly, et al. *Phys. Rev. Lett.* 111, 080502 (2013)

"Superconducting quantum circuits at the surface code threshold for fault tolerance", Rami Barends, Julian Kelly, et al. *Nature* 508, 500-503 (2014)

"Optimal quantum control using randomized benchmarking", Julian Kelly, Rami Barends, et al. *Phys. Rev. Lett.* 112, 240504 (2014)

"Rolling quantum dice with a superconducting qubit", Rami Barends, Julian Kelly, et al. *Phys. Rev. A* 90, 030303(R) (2014)

"Overcoming correlated noise in quantum systems: How mediocre clocks make good qubits" Peter O'Malley, Julian Kelly, Rami Barends, et al. *submitted* (2014)

"State preservation by repetitive error detection in a superconducting quantum circuit" Julian Kelly, Rami Barends, Austin Fowler, et al. *Nature* 519, 66-69 (2015)

Abstract

Fault-tolerant superconducting qubits

by

Julian S. Kelly

For quantum computing to become viable, the inherently fallible nature of qubits must be overcome with quantum error correction (QEC). QEC requires coherent qubits, in a configuration compatible with a given QEC scheme, and quantum logic operations with sufficiently low error. In this thesis, we perform a series of targeted experiments to achieve these goals on a path toward realizing the surface code QEC scheme. We first develop the Xmon variant of the transmon qubit, a highly coherent, planar, and frequency tunable superconducting qubit. With coherence demonstrated, we build an array of five Xmon qubits in a configuration compatible with the surface code, and demonstrate quantum logic operations with sufficiently low error to employ the surface code. These logic gates are characterized with randomized benchmarking, a protocol for determining gate error. We find applications of randomized benchmarking beyond the intended use in gate optimization and decoherence characterization, in addition to exploring the fundamental assumptions of randomized benchmarking. Lastly, we build a nine-qubit Xmon transmon array and demonstrate correction of environmental bit-flip errors in a precursor to the surface code.

Contents

Chapter 1

Quantum computing and error correction

1.1 Introduction

Information processing technology has revolutionized science, how we conduct our personal lives, and what drives our economy. The pervasiveness of such technology is astounding; numerically solving a differential equation, checking your email, and your job all critically rely on systems that accurately and efficiently process information. The need to solve complex problems is massive: countries and corporations all over the world have spent hundreds of millions on constructing supercomputing clusters to tackle the hardest of problems, and organize vast quantities of data.

Yet, even with these computers that can have hundreds of thousands of processing cores, 10^{14} transistors, and consume megawatts of energy, many problems are still out of reach. In fact, there are some tasks that computer scientists are pessimistic about *ever* being able to solve with modern computer technology. With relative ease, one can construct a problem that would

take more energy than exists in earth's fossil fuels, and longer than the age of the universe to solve on modern hardware. Thus, there is great interest in finding unconventional approaches to solving hard problems.

1.2 Quantum computing

Large quantum mechanical systems are notoriously complex as the degrees of freedom scale exponentially with system size. It is for this exact reason that quantum computing is so intriguing: the natural evolution of quantum systems vastly outpaces what can be simulated with modern computing. The field of quantum computing is on a journey to harness this power.

A quantum computer is a device that seeks to use the fundamental physics of quantum mechanics, such as superposition and entanglement, for computation. They do so through manipulations of quantum bits (qubits), the quantum analog to classical bits. It is possible to find new algorithms run with qubits that can solve problems with significantly fewer resources than the best known classical algorithms. One such example is Shor's algorithm, which can be used to efficiently factor large numbers into their primes, with applications in cryptography. Although other unequivocal examples of quantum speedup remain somewhat elusive, there is great hope that new applications will become readily apparent as the technology matures.

1.2.1 Superposition and entanglement

Classical bits can take on the values of 0 or 1, whereas qubits can be $|0\rangle$, $|1\rangle$, or a quantum *superposition* of both. The state of a qubit $|\psi\rangle$ can be parameterized as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1.1}$$

$$= \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle \tag{1.2}$$

where the complex amplitudes α and β are constrained by $|\alpha|^2 + |\beta|^2 = 1$, as the probability of a qubit being in $|0\rangle$ or $|1\rangle$ is 1. This has a useful geometrical interpretation, where θ and ϕ are the angles specifying the direction of a unit vector in three dimensions: this is called the Bloch sphere interpretation.

The real magic of quantum computing comes in when multiple qubits are involved. A two qubit system has four basis states instead of two: $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$. These additional states, come with their own complex amplitudes: $|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle$, with $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$. The complexity of a qubit system scales exponentially; adding a single qubit to a quantum computer doubles the number of basis states and complex amplitudes available for computation. Thus, an n qubit quantum computer has 2^n bases states and amplitudes available for computation. It is this scaling that we hope to tap into to solve problems.

1.2.2 Errors are everything

Quantum computing has greater intrinsic challenges than classical computing. This fundamentally comes down to how information is stored. Suppose we store a bit of information as a voltage in a classical circuit, we define a 0 as 0 V, and 1 as 5 V. As any experimental physicist is well aware, you will never actually measure exactly 0 or 5 V in the lab. A wonderful feature of classical logic is that it is forgiving; if we measure 0.1 V we call that a 0, and if we measure 4.95 V we will call that a 1. This leniency, combined with feedback that pushes voltages back to 0 or 5 V, makes information stored this way very robust.

Qubits however, are a bit more tricky. If we encode our qubit in a state with angle $\theta + \epsilon$ instead of θ, the qubit is in a distinctly different state than we intended. This is because quantum information is stored in continuous amplitudes and phases, whereas classical information is discrete. If we then measure our qubit, we will have probability $\cos((\theta + \epsilon)/2)^2$ to get the $|0\rangle$ state, a physically measurable error.

An even bigger problem is that of quantum coherence. Quantum states are extremely fragile; any interaction of a qubit with its environment can cause decoherence, or the loss of information stored in the qubit. Quantum decoherence is broken down into two effects: energy relaxation and dephasing, with corresponding characteristic timescales T_1 and T_2, known as coherence times. In systems such as superconducting qubits, the $|1\rangle$ state naturally relaxes into the $|0\rangle$ state. The effect of energy relaxation in the Bloch sphere is a gentle pull towards the north pole of the sphere, or a corruption of θ. Dephasing is the blurring of the relative phase ϕ between $|0\rangle$ and $|1\rangle$; in the Bloch sphere representation this is a shrinking of the x and y

components of the Bloch vector. Thus, information in a quantum system has a finite lifetime, and is much more volatile than in a classical system. Coherence is one of the fundamental challenges in quantum computing: in order to use a qubit for computation, you need to be able to control it and interface it with other qubits, yet anything else the qubit is externally coupled to may corrupt it.

This leaves us with an empirical question: does nature allow useful computation using quantum mechanics given the existence of decoherence, and the continuous nature of quantum information?

1.3 Quantum error correction

1.3.1 Motivation

Theorists estimate that a useful quantum algorithm would require on the order of 10^{20} physical quantum logic operations. This number, although large, is typical for classical processing tasks. In order to successfully execute a quantum algorithm, each qubit would need to be able to implement quantum logic operations with error rates on the 10^{-20} level. This is challenging for two reasons. First, the raw coherence times must be long enough such that a qubit can perform the required number of logic operations. Secondly, control must be precise enough such that qubit manipulations generate the desired quantum states without becoming inundated with accumulation of error.

How feasible is this? Scientists in quantum computing have been diligently working the last

20 years to achieve low error rates in a variety of different platforms. To date, the lowest error rates (incorporating control and decoherence error) for quantum logic operations in any system are 10^{-6} for single qubit operations in trapped ions [**?**], and 10^{-2} for two-qubit operations in a variety of platforms [**?** **?** **?** **?**]. However, these record numbers must also translate out of simple experiments and into more complex systems with many qubits and control wires.

Optimistically, errors are at least 10^{15} orders of magnitude away from where they need to be for useful computation. Why is this so challenging? For superconducting qubits, logic operations take on the order of 10 ns, and coherence times are on the order of 10 μs, meaning that one can perform roughly 10^3 logic operations within a coherence time. Additionally, we control our qubits with waveforms generated from a 14-bit digital to analog converted (DAC). As a simple example, suppose physical qubits are perfect and we only consider error from control. Given a set of logical operations to rotate our qubit state between 0 and π, this means we have an angular resolution of $\pi/2^{14} \sim 10^{-4}$ radians. Radial accuracy must be roughly 10^{-10} to achieve required fidelities, error $\sim \phi^2/4$. Improving either of these numbers to the required error rates is practically impossible. On top of this, these numbers would have to be achieved across a large number of qubits which have other imperfections and nonidealities. This is where quantum error correction comes in.

1.3.2 Fault-tolerance

In classical computing, hardware is extremely reliable, but imperfect. To account for this, classical computing schemes are made to be fault-tolerant, where infrequent errors can be tolerated

without corrupting the output of an algorithm. This is done by introducing redundancy, where additional pieces of hardware or copies of information allow for infrequent errors to be detected and ignored.

Similar techniques can be used for quantum computing, where many physical qubits working together can be used to construct what is called a *logical* qubit, with significantly reduced error rate. A number of different fault-tolerant quantum computing schemes exist, which permit computations using logical qubits in the presence of a small number of errors on the physical qubits which compose them [**?** **?** **?**]. These schemes must be carefully constructed to permit the detection of errors without destroying quantum information.

1.3.3 Measuring errors

Detecting errors without destroying a quantum state requires care. We saw previously that we could detect an error ϵ by directly measuring the qubit state. However, we know that measurement of our qubit will project it into the $|0\rangle$ or $|1\rangle$ state, destroying quantum information α and β. However, by encoding a qubit state onto many physical qubits with extra degrees of freedom, we can construct measurements that do not perturb our quantum information. We will elucidate this with an example.

1.3.4 Five qubit error detection code

Suppose we take a single qubit state and map it onto three physical qubits:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \mapsto \alpha|000\rangle + \beta|111\rangle \tag{1.3}$$

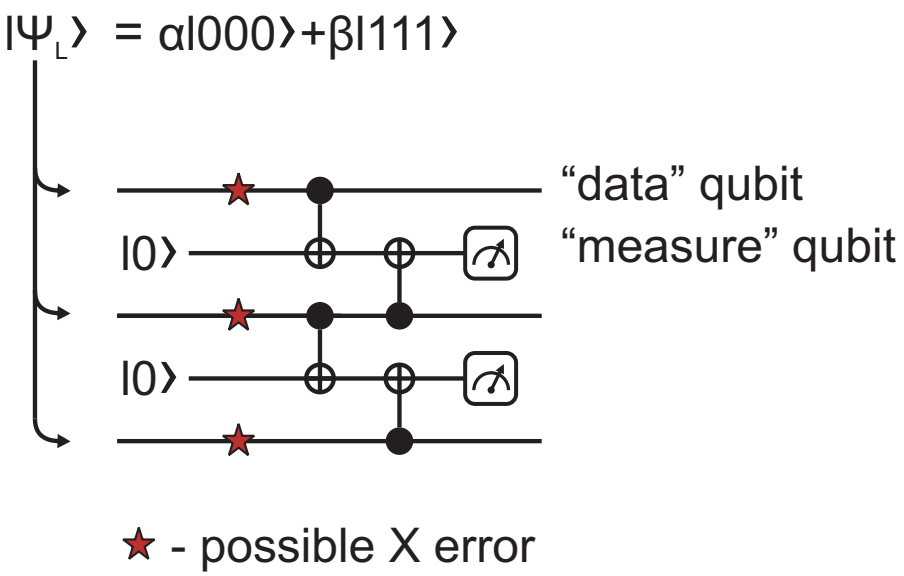


Figure 1.1: **Detecting bit flips with ancilla qubits.** Quantum information (amplitudes α and β) is encoded onto a three qubit state. Through the use of two ancilla qubits performing CNOT gates, we can extract which data qubit had a bit flip error.

Now, suppose we introduce a bit flip ($\hat{X}$) error (mapping $|0\rangle \mapsto |1\rangle$ and vice versa) onto a single qubit, or no error at all. There are four outcomes:

$$|\psi\rangle_0 = \alpha|000\rangle + \beta|111\rangle \tag{1.4a}$$

$$|\psi\rangle_1 = \alpha|100\rangle + \beta|011\rangle \tag{1.4b}$$

$$|\psi\rangle_2 = \alpha|010\rangle + \beta|101\rangle \tag{1.4c}$$

$$|\psi\rangle_3 = \alpha|001\rangle + \beta|110\rangle \tag{1.4d}$$

First, notice that our quantum information, stored in α and β still exists. In order to continue to use our state for computation, we need to determine what type of error has occurred. We know that direct measurement will collapse the state, so we need to be more clever.

We introduce two ancilla "measure" qubits to probe the "data" qubits that store our quantum information. Consider the circuit in figure 1.1. The controlled-NOT (CNOT) gates flip the state of the measure qubit, conditional on the state of the data qubit. Each measure qubit probes two data qubits, equivalent to measuring the $\hat{Z}\hat{Z}$ operator. If the data qubits are in $|00\rangle$ or $|11\rangle$, the measure qubit will be in the $|0\rangle$ state. If the data qubits are in the $|01\rangle$ or $|10\rangle$ state, the measure qubit will be in the $|1\rangle$ state. Thus, by looking at the output state of the measure qubits ($|00\rangle$, $|01\rangle$, $|10\rangle$, or $|11\rangle$), we can determine which data qubit (if any) suffered a bit flip error. Importantly, the logical state we have constructed has not been perturbed by our measurement; we have not collapsed α and β. This is because the states $|\Psi\rangle_{0-4}$ are eigenstates of the $\hat{Z}_1\hat{Z}_2$ and $\hat{Z}_2\hat{Z}_3$ operators.

1.3.5 Coherent errors

This five qubit example is pedagogically interesting, but appears impractical. It is difficult to believe that nature would hand us a perfect bit flip to detect with this method. As in the previous section, we were worried about small angle errors ϵ away from the ideal state. We will now work through the previous example in the case of a small coherent error ϵ, and find that the measuring the measure qubits will either force the system to have an $\hat{X}$ error or no error at all.

Suppose instead of a single bit flip, we apply an X rotation. We first initialize the data qubits into the state we wish to preserve, and the measure qubits into the ground state:

$$|\Psi_{\text{initial}}\rangle = \gamma(\alpha|000\rangle + \beta|111\rangle) \otimes |00\rangle \tag{1.5}$$

We then apply an X rotation to the first data qubit $|0\rangle \mapsto \gamma|0\rangle + \delta|1\rangle$:

$$|\Psi_{\mathrm{X}}\rangle = (\gamma(\alpha|000\rangle + \beta|111\rangle) + \delta(\alpha|100\rangle + \beta|011\rangle)) \otimes |00\rangle \tag{1.6}$$

Finally, we apply the CNOT gates and map the bit error onto the ancilla qubits:

$$|\Psi_{\text{final}}\rangle = (\alpha|000\rangle + \beta|111\rangle) \otimes \gamma|00\rangle + (\alpha|100\rangle + \beta|011\rangle) \otimes \delta|10\rangle \tag{1.7}$$

If we then measure the measure qubits, we will find they were in the $|00\rangle$ state with probability $|\gamma|^2$ or the $|10\rangle$ state with probability $|\delta|^2$. The act of measurement will project the the continuous case onto either the case of no error, or the case of a perfect bit flip error.

1.3.6 X and Z errors

The circuit that we have outlined can only be used to detect $\hat{X}$ errors. In order to make a system that is fully fault-tolerant, we must protect our qubits from arbitrary rotation errors, which can be decomposed into $\hat{X}$, $\hat{Y}$, and $\hat{Z}$. $\hat{Y}$ errors can be decomposed further, $\hat{Y}$=$i\hat{Z}\hat{X}$. Given this, detecting $\hat{X}$ and $\hat{Z}$ errors is sufficient.

Simultaneously detecting $\hat{X}$ and $\hat{Z}$ errors is nontrivial. This can be understood by the commutation relation of measurement operators. We stated previously that measuring $\hat{X}$ errors is equivalent to measuring $\hat{Z}$ operators, as the $|0\rangle$ and $|1\rangle$ states are eigenstates of $\hat{Z}$. We would also like to simultaneously measure the $\hat{X}$ operator to detect $\hat{Z}$ errors. However, $[\hat{X}, \hat{Z}] \neq 0$, meaning that we cannot find a simultaneous eigenstate of both operators. Sequential mea-

surements of $\hat{X}$ and $\hat{Z}$ operators will have back action that perturbs our quantum information. Alternatively, this can be understood by seeing that either a $\hat{X}$ or $\hat{Z}$ measurement will fully constrain the quantum system to be in an eigenstate of that operator; there are no remaining degrees of freedom to store information in after such a measurement. To get around this, we can construct multi-qubit operators, $[\hat{X}_1\hat{X}_2, \hat{Z}_1\hat{Z}_2] = 0$. These operators have simultaneous eigenstates, the Bell states.

Take the example of Fig 1.2, with two data qubits and two measure qubits. As there are two data qubits, there are four basis states and amplitudes. The measure qubits apply the $\hat{X}_1\hat{X}_2$ and $\hat{Z}_1\hat{Z}_2$ operators. The four possible outcomes if these operators will fully constrain the system to one of four possible Bell eigenstates of said operators. Although there are not enough degrees of freedom to store quantum information in one Bell state, this simple example demonstrates that $\hat{X}$ and $\hat{Z}$ operators can be simultaneously measured, without perturbing an entangled state. In order to perform computations, additional data qubits and thus degrees of freedom must be introduced. The specifics of how this is done depends on error correction scheme.

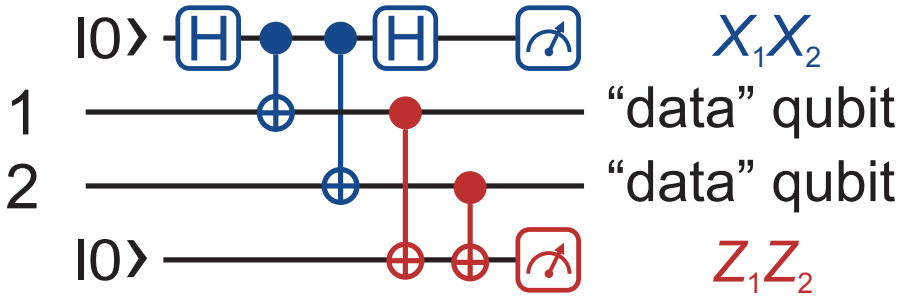


Figure 1.2: **Simultaneous $\hat{X}$ and $\hat{Z}$ operators.**

1.3.7 Decoherence errors

We have described the case where we have a coherent error that rotates our qubit state, but as mentioned previously, decoherence can cause errors that are incoherent. Decoherence errors can be expanded in terms the identity and $\hat{X}$, $\hat{Y}$ and $\hat{Z}$ errors. The example in Fig.1.2 will force a decoherence error to manifest as an identity, $\hat{X}$, $\hat{Z}$ or $\hat{Y} = i\hat{Z}\hat{X}$ error. After projection, the system will again be forced into a Bell state.

1.3.8 Measure qubit errors

So far, we have only considered the case that we have errors on data qubits. Unfortunately, measure qubits can also have errors. To combat this, we can repeat the error detection algorithm multiple times. Suppose two data qubits are initialized into the $|00\rangle$ state, and a measure qubit performing the associated $\hat{Z}_1\hat{Z}_2$ operation is initialized into $|0\rangle$. A data qubit error will change the state of the data qubits to $|01\rangle$ or $|10\rangle$, and upon performing error detection circuit, the measure qubit will end in the $|1\rangle$ state. If the error detection circuit is repeated, the measure qubit will flip state in the subsequent round, ending in $|0\rangle$. Alternatively, an error on a measure qubit will leave the measure qubit in the $|1\rangle$ state at the end of the first round, but leave the data qubits in the $|00\rangle$ state. Subsequent error detections will *not* change the state of the measure qubit, and it will remain in the $|1\rangle$ state. Interestingly, after the first round it is impossible to distinguish between a measure qubit error and data qubit error, but subsequent detections allow us to distinguish the cases.

1.3.9 Threshold

Up to this point, we have shown that we can identify different possible errors on data and measure qubits, assuming errors are sufficiently sparse. However, in a realistic system we expect to see multiple errors which all must be individually identified to prevent algorithmic failure. This is the basic concept behind the quantum threshold theorem: if errors are sparse they can be individually identified, as in the simple examples we have worked through up to this point. Additionally, in this limit larger systems will be even more resilient to error. However, if the density of errors increases beyond some threshold, errors cannot be reliably identified. Mistakes in doing so will lead to logical failure. Given a system below threshold, it would be possible to achieve error rates on the 10^{-20} level by increasing the number of physical qubits per logical qubit, making it less likely a large number of errors will occur simultaneously.

1.4 Conclusion and outlook

Quantum computing looks to be a promising candidate to extend computational power beyond classical computers. However, various practical limitations in control and qubit coherence introduce error. We have outlined some of the fundamental building blocks for detecting quantum errors without destroying quantum entanglement and superpositions: multi-qubit $\hat{X}$ and $\hat{Z}$ operators. By constructing an architecture that implements these operators, we can build logical qubits from many physical qubits which are fault-tolerant to errors. Given sufficiently reliable operations, it is possible to suppress errors to the levels required for useful computation.

The rest of this thesis will be concerned with experimentally implementing practical error

correction. There are a number of challenges ahead: how do we build an architecture than can implement large scale quantum error correction? How do we design a physical qubit system compatible with this architecture? Can we really achieve the required logic operation fidelities, and show quantum error correction providing a benefit? I hope to answer these questions by the end of this thesis.

Chapter 2

Superconducting qubits and the surface code

2.1 Candidate quantum computing systems

There are many candidate physical systems for a quantum computer; in principle any two quantum states can make be used as a qubit. Implementations that revolve around isolating small quantum objects are popular: single electrons in quantum dots [**?** **?**], electronic states of a trapped ion [**?** **?** **?** **?** **?** **?** **?**], and intra-molecular spin interactions in nuclear magnetic resonance [**?** **?** **?** **?**] have all had varying degrees of success at showing fundamentals of quantum computing. Superconducting circuits are unique in comparison to these systems; superconductivity collapses a metal's Fermi sea of electrons into a single, macroscopic quantum state through the creation of Cooper pairs.

The characteristic differences in these systems highlights a tension in all quantum computing platforms, the careful balance of coherence, connectivity and control; all of which are essential elements for computation. In systems of quantum dots, trapped ions, and NMR, qubit lifetimes can be extraordinarily long: coherence times on the order of seconds [**?**], minutes [**?**], and hours [**?**] have been demonstrated. However, these coherence times are achieved by successfully isolating a quantum system from the environment, introducing new challenges: how do you control and connect extremely isolated qubits?

The electronic states of trapped ions naturally have extremely weak interactions, thus multi-qubit logic can be difficult to achieve. This has been overcome by using atomic motional modes as a medium to couple electronic states in different atoms together, but introduces a new set of challenges. Interacting ions are generally placed into the same trap, where Coulomb forces hold them mere microns apart (see Fig. 2.1b) which can make addressability difficult. This can be overcome with a variety of techniques, such as using field gradients to break degeneracy allowing frequency addressability with microwaves, using tightly focused lasers to avoid spacial overlap, and actively moving ions apart to reduce crosstalk. All of these techniques are under active investigation.

In NMR, the only non-neglible interactions are between atoms within the same molecule. As these molecules are quite small, interfacing to qubits is done with global fields and frequency multiplexing. Additionally, larger qubit systems can only be realized with increasingly complex molecules. Theses challenges are considered to be insurmountable, an example of a system of coherence but without proper balance with connectivity and control.

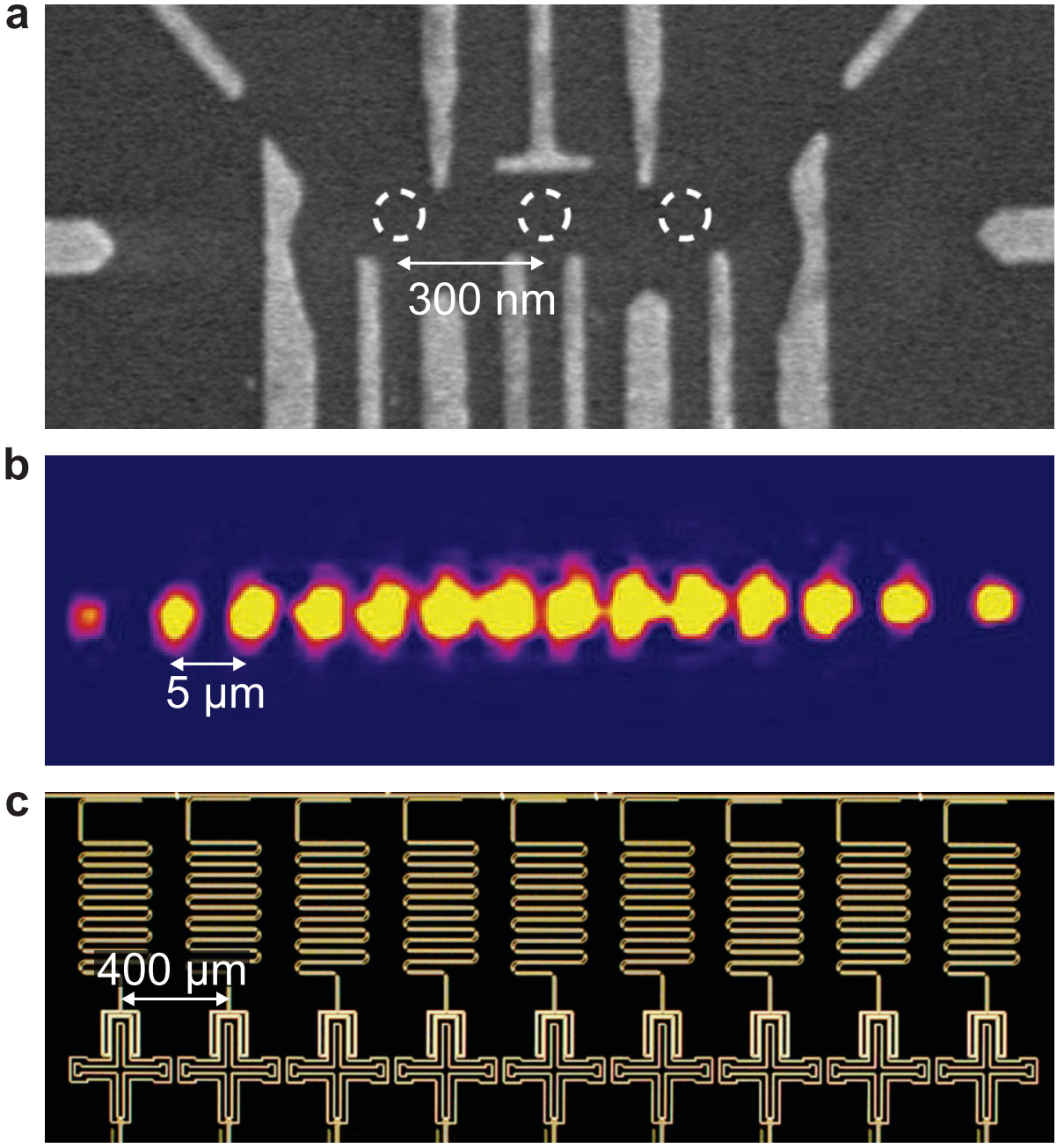


Figure 2.1: **Candidate quantum computing systems.** (a) Image from TU Delft. Electron microscope image of triple electron quantum dot. Dashed lines indicate placement of isolated electrons, with spacing of 300 nm. (b) Image from R. Blatt group, Innsbruck, Austria. Image of 14 trapped ions in a linear Paul trap, with spacing of roughly 5 μm. (c) Nine superconducting Xmon transmon qubits, spaced 400 μm apart.

2.2 Superconducting circuits

Superconducting circuits lie on the opposite end of the spectrum: a macroscopic quantum state makes coupling and control easy, as the properties and shape of the state can be engineered, see Fig 2.1c. For the same reason, coherence becomes an obstacle. Isolating a macroscopic state from unwanted interactions with its environment is challenging. Historically, superconducting qubits have been hampered by short coherence times. However, if this can be overcome, the other aspects of superconducting circuits hold incredible promise. Modern electronics are based off of circuits, and tapping into this technology comes with all of the advantages.

One of the great promises of superconducting circuits is the natural compatibility with the cleanroom technology used for semiconductor circuits. This allows superconducting circuits to directly benefit from highly reliable and scalable microfabrication technology. Superconducting circuits are generally of modest size (~ 100 μm) and frequency (few $\sim$ GHz), which is advantageous. When qubits are of this size, routing μm scale control wiring is simplified dramatically. Use of GHz electronics is widespread due to its usefulness in radio frequency transmission; these highly engineered systems have become remarkably cheap and reliable, in contrast to, for example, laser based systems. The circuit nature of superconducting qubits allows coupling through simple mechanisms such as capacitance and inductance, all of which can be optimized through microwave engineering methods. Additionally, classical circuit elements such as linear resonators [**?**] and filters [**? ?**] can be built natively into qubit systems.

2.2.1 The Josephson junction

The heart of superconducting quantum circuits is the Josephson junction. This superconducting phenomena manifests as a non-linear inductor, which can have a variety of uses in superconducting systems. This non-linearity is the critical element that breaks the degeneracy of linear circuits. Josephson junctions can be used to create anharmonic oscillators, in which different state transitions can be individually addressed with a different frequencies. All varieties of superconducting qubits are based off of non-linear LC oscillators, where dissimilarities in the energy level spacing allow for two quantum states to be individually addressed as a qubit. Junction uses go beyond static non-linearity: two junctions in parallel create a superconducting quantum interference device (SQUID), effectively a tuneable junction that can be adjusted near-instantaneously with a flux bias. This powerful element allows qubit interactions to be turned on and off quickly by adjusting qubit frequencies [**?**] or directly modifying the inductive coupling between qubits through use of a junction [**? ? ?**]. Last but not least, junction non-linearity can be used to faithfully amplify extremely low power signals, adding only the minimum noise required by quantum mechanics [**?**]. These quantum-limited amplifiers have revolutionized readout of superconducting qubits by dramatically increasing the signal to noise ratio in the dispersive readout scheme [**? ? ?**].

2.2.2 Coherence

These wonderful properties of superconducting circuits only become relevant if coherence times are significantly longer than the length logic operations. Logical operations typically

take tens of nanoseconds, given qubit nonlinearity and couplings in the MHz range. Current estimates on the requirements for error correction mandate that coherence times reach at least tens of microseconds. To illustrate this challenge, I note that one of the first superconducting qubit devices had a coherence time of 2 nanoseconds [**?**].

What limits the coherence times of superconducting circuits? Over a decade of research has identified a series of culprits: circuitry, magnetic vortices, quasiparticles, and materials loss. The simplest is control circuitry. Any control wire that couples into the qubit circuit for manipulations also provides a path for energy stored in the circuit to leave [**?**]. Empirically, it is possible to engineer control circuitry that allows both for qubit manipulations and acceptable coherence times. Magnetic vortices are induced when a magnetic field is strong enough to permeate the superconductor creating a normal metal region that can lead to inductive losses. This is avoided by protecting the superconducting circuitry in a magnetic shield, and ensuring that metallic components nearby are non-magnetic. Quasiparticles result from high energy processes, such as infrared radiation incident on the superconductor, which break Cooper pairs. These quasiparticles can induce decoherence in the qubit circuit if they happen to tunnel through the junction [**?**]. The effect of quasiparticles can be mitigated by introducing material that absorbs infrared radiation on the control lines and in the sample area [**?**]. As these processes have to do with circuit design or the sample environment, rather than specifics of how the samples are fabricated, they are straightforward to mitigate.

2.2.3 High quality materials

The performance of a superconducting circuit is dramatically effected by the materials that compose it, thus the recent success of superconducting circuits can largely be attributed to improvements in materials research. Although bulk superconductors are intrinsically lossless, nearby dielectrics must be of high quality. Electronic circuits are fabricated by patterning layers of metal onto, in our case, a non-conductive dielectric substrate. However, dielectrics can damp electromagnetic fields that permeate them, introducing decoherence in qubit systems [**?**]. Improving the coherence properties of materials can be particularly tricky. Different materials can have dramatically varied processing requirements, making controlled experiments difficult to carry out. However, due to persistent, careful experimentation, advances in materials [**?**] and knowledge of the previously outlined decoherence mechanisms [**?**] have allowed integrated circuit lifetimes to improve from 2 ns [**?**] to 10's of μs or more [**? ?**] in the last 15 years.

There is an alternative approach to increasing circuit coherence without improving the underlying materials, which is to change the circuit geometry to move electromagnetic fields out of lossy materials and into lossless vacuum. This is the approach taken by the 3D transmon [**?**]. In this geometry, a circuit is suspended in a three-dimensional (3D) cavity. At the time, this qubit achieved dramatically improved coherence times over the current state of the art, up to about 100 microseconds. This experiment highlights the trade offs in quantum computing platforms; the 3D transmon benefits from extended coherence, yet the isolated cavity architecture introduces challenges in connectivity and control. Approaches for reliable connectivity

and robust control are still being heavily investigated today for 3D systems.

Interestingly, systems which achieve coherence through isolation place a higher emphasis on materials research as they strive to achieve higher levels of connectivity and control. Recently, microchips designed to trap ions have become a focus of research [**?**]. These devices bring the sophistication of integrated circuits to ion trap qubits. However in these systems, ion motional modes can be excited due to electromagnetic fluctuations from materials on the chip surface [**?** **?**]. In spin qubits, nuclear spins surrounding isolated electrons can be a a source of frequency fluctuation and thus dephasing [**?**]. In 3D transmons, connecting multiple qubits and cavities degrades coherence due to details of how cavities and qubits are constructed and interconnected. Thus, all architectures face challenges in materials.

We have found that coherence, connectivity and control must all be carefully considered and balanced when analyzing a quantum computing candidate system. In the same way, these must be considered when choosing an implementation of error correction. If quantum computing becomes viable, it will be due to a physical system that is well matched to an error correction protocol.

2.3 The surface code

In this section, we introduce the surface code quantum error correction (QEC) scheme, how it works, and describe the natural synergy between the surface code and superconducting qubits.

2.3.1 Operation and connectivity

One of the most important considerations when physically implementing a computer is that of connectivity. The basic components for QEC are outlined in Chapter 9, but a physical architecture is needed to realize them. Geometry of qubit locations, control wiring, and feasibility of fabrication all must be taken into consideration. For example, the ability to interact arbitrary pairs of qubits can be very powerful on paper, implementing a mechanism that allows this while maintaining operation fidelity, coherence, etc. is impractical. Given the complexity of realizing small systems of reliable, low error qubits, a simple architecture is desirable for engineering and scaling purposes.

The surface code provides such a simple architecture. The surface code acts on a two-dimensional array of qubits, where the purpose of each qubit alternates in a checkerboard pattern, see Fig. 2.2a. Data qubits, corresponding to the open circles, store quantum information while measure qubits, corresponding to the filled circles, detect errors on data qubits. Thus, each data (measure) qubit interacts with only its four nearest neighbor measure (data) qubits. Measure qubits come in two varieties, one to detect $\hat{X}$ and one to detect $\hat{Z}$ errors (Fig 2.2b and c). Each data qubit has two measure-X and two of the measure-Z neighbors, such that both $\hat{X}$ and $\hat{Z}$ errors can be detected. A square lattice with N data qubits and $N-1$ measure qubits has exactly two degrees of freedom which constitute the logical qubit. Through repetitive $\hat{X}$ and $\hat{Z}$ measurements, errors can be detected on data qubits and distinguished from errors on measure qubits, thus the surface code is fully fault-tolerant. Only single qubit gates, CNOT gates and individual measurement is required for error correction. The simple geometries and

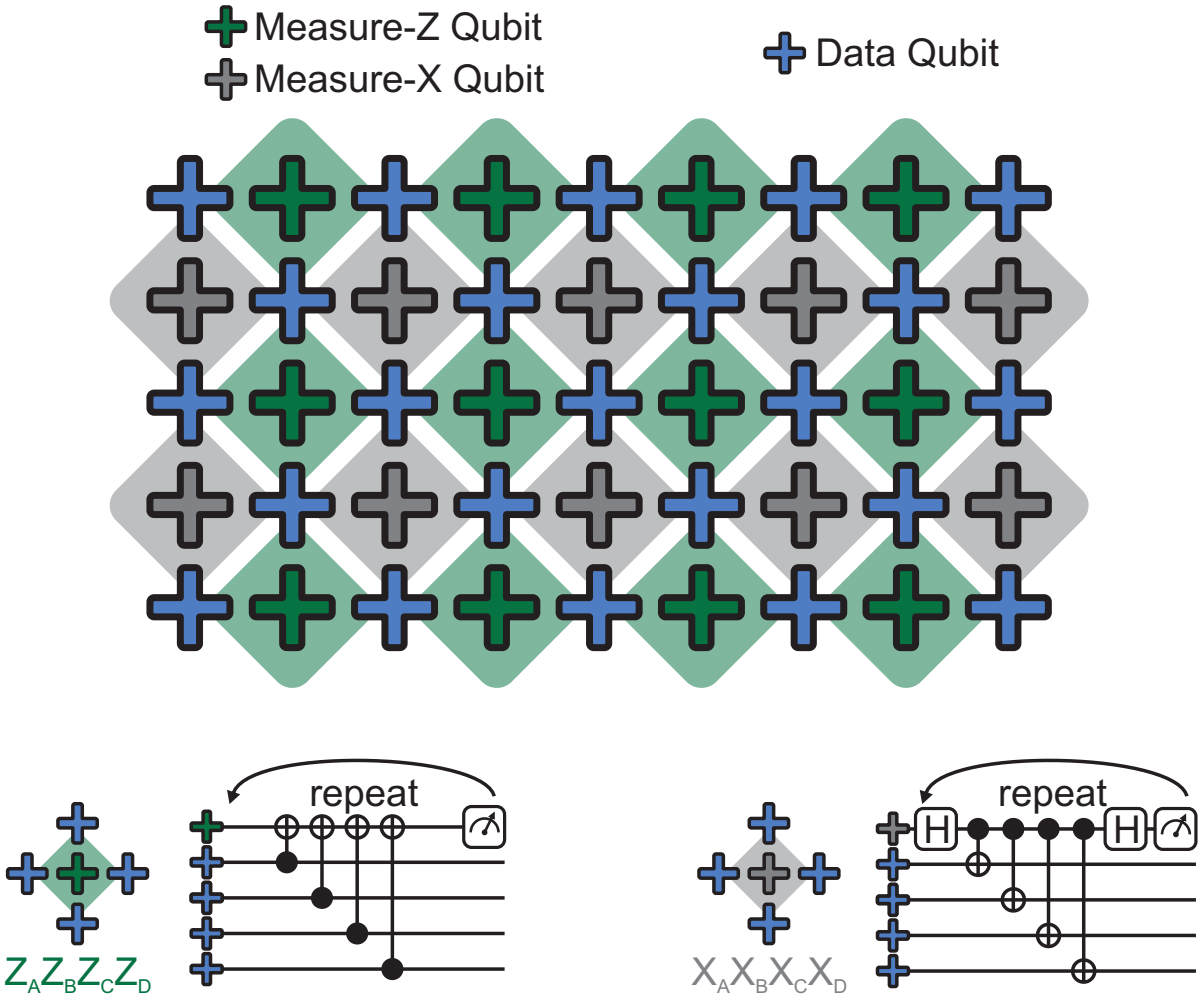


Figure 2.2: **The surface code QEC achitecture.** The surface code is a two-dimensional array of qubits in an alternating scheme. Data qubits (blue crosses) store data, while measure qubits (grey and green crosses) detect errors on data qubits. Measure qubits come in two varieties, measure-X and measure-Z, which detect $\hat{X}$ and $\hat{Z}$ errors depending on the polarity of CNOT gates. Detection cycles are repeated, and are used to distinguish between data and measure errors.

gate requirements make the surface code of great practical interest.

Due to the simple configuration and gate short sequence, the error threshold is roughly 1% per operation, significantly more forgiving than other QEC schemes such as the Steane code with rates as demanding as 10^{-4}. This low error rate is achieved by moving complexity out of the quantum hardware and into a classical algorithm. This algorithm uses the output of multi-qubit $\hat{X}$ and $\hat{Z}$ measurements to detect physical errors. In the surface code, errors need only be detected rather than corrected, reducing demand on the hardware. These detected errors can be efficiently commuted through logical Clifford operations, and be classically corrected for at the end of computation. Thus, dynamical feedback during computation is only required to implement the non-Clifford T gate at the logical qubit level. These features greatly relax the requirements for quantum hardware, which historically have kept practical error correction out of reach.

Superconducting circuits are naturally compatible with the surface code QEC scheme. The planar, two-dimensional qubit geometry in the surface code lends itself to layer-based microfabrication. Nearest neighbor interactions can be simply realized with capacitive or inductive coupling. Readout and control requiring wiring, and as modest qubit sizes reduce circuit density to manageable levels, routing control wiring for many qubits is likely a surmountable task. As the error threshold is forgiving, the relatively modest coherence of superconducting circuits may not pose an impediment. Thus, the combination of the surface code with superconducting qubits promises to be a strong candidate for a quantum computer.

2.3.2 Realizing the surface code

The fundamental requirements for a superconducting-circuit based surface code quantum computer have now been outlined. At this point, I discuss a series of experiments to realize the hardware requirements for a superconducting circuit-based surface code quantum computer. We begin by building a highly coherent qubit compatible with the surface code architecture, the Xmon (Chapter 3), move to implementing high fidelity gates (Chapters 4-7) and measurement (Ref. [**?**]), and finally demonstrate basic surface code operations (Chapter 8).

Chapter 3

Coherent Josephson qubit suitable for scalable quantum integrated circuits[1]

3.1 Introduction

One of the outstanding challenges in building a quantum computer is to balance coherence, connectivity and control in the qubits. Superconductivity provides an appealing platform because it allows for scalability: the conduction electrons condense into a macroscopic quantum state, and large quantum integrated circuits can be made with many elements having individual control lines. However, quantum coherence in superconducting circuits has proven to be very delicate, as it is easily disturbed by material defects, electron system excitations, and radiative coupling to external wiring [**? ? ? ? ? ? ? ? ?**]. To minimize these and other effects, many

[1]The following was published as: "Coherent Josephson qubit suitable for scalable quantum integrated circuits", Rami Barends, Julian Kelly, et al. *Phys. Rev. Lett.* 111, 080502 (2013).

groups have recently begun embedding qubits in three-dimensional superconducting cavities. These 3D qubits show high coherence, with energy relaxation times in 3D transmon qubits between 30 and 140 μs [**? ?**].

In this chapter, we demonstrate a new design for a fully planar superconducting qubit, based on the planar transmon [**? ?**], with energy coherence times in excess of 40 μs. Our approach balances coherence, connectivity, as well as fast control. The qubits are frequency-tunable, which allows the implementation of fast two-qubit gates: a controlled-Z gate [**? ? ?**] can then be implemented with high fidelity in 25 ns [**?**]. With the coherence time exceeding single and two-qubit gates times by three orders of magnitude, we believe that our device provides a key ingredient for implementing a surface code quantum computer [**?**].

We also identify an incoherent decoherence mechanism, arising from a sparse bath of weakly coupled defects. This incoherent regime is made accessible by the long coherence of our qubits. We explore this physics by visualizing these defects in the measured quantum time-resolved spectroscopy, by varying the qubit geometry, and by a model analysis. These defects give rise to frequency-dependent variations in the lifetime; our results may also explain the variations observed in lifetimes of 3D transmon qubits.

3.2 Design perspective

The Xmon transmon developed here is purpose built for implementing the surface code. When we began this project in mid 2011, our lab exclusively used the phase qubit. Advantages of the phase qubit include a simple readout and coupling scheme, but coherence had always been a

large challenge. T_1 times of typically 0.5 μs were limited by dielectric losses due to a parallel plate capacitor [**?**], and a natural flux sensitivity meant that $1/f$ flux noise – present in all SQUID based qubits – limited T_2 to 0.1-0.2 μs [**?**]. At this point, transmon qubits were becoming quite popular [**?**]. This qubit has the advantage of naturally using small planar capacitors where low-loss substrates and perfect vacuum were used as a dielectric. Given that the surface code only requires nearest neighbor interactions, the small capacitance made the transmon a natural choice to achieve this modest level of connectivity. Additionally, the transmon is less flux sensitive than the phase qubit, thus T_2 times could be dramatically longer with the same levels of flux noise. Typically, planar transmon qubits had T_1 of 1-3 μs, and T_2 of roughly 0.5-1 μs, depending on bias point [**?**].

In early 2011, the three-dimensional (3D) transmon was introduced [**?**]. This experiment took a transmon qubit fabricated on an empty substrate, and suspended it in a 3D microwave cavity. Additionally, this qubit had a single junction rather than a SQUID loop, removing both frequency tunability and the effects of $1/f$ flux noise. This architecture dramatically improved coherence times given the same materials as planar transmons, the original publication showed $T_1 = 20 - 60$ μs, $T_2 = 6 - 20$ μs. This experiment demonstrated two important things: planar transmons were limited by dielectric loss in the capacitor as moving E-field lines away from the surface and into vacuum improved coherence, and that the Al/AlO_x/Al Josephson junction was intrinsically low loss. Additionally, this geometry greatly simplified the microwave environment from an integrated circuit, making it easier to probe the raw coherence of the circuit.

At this point, much of the field began researching 3D architectures as extended coherence

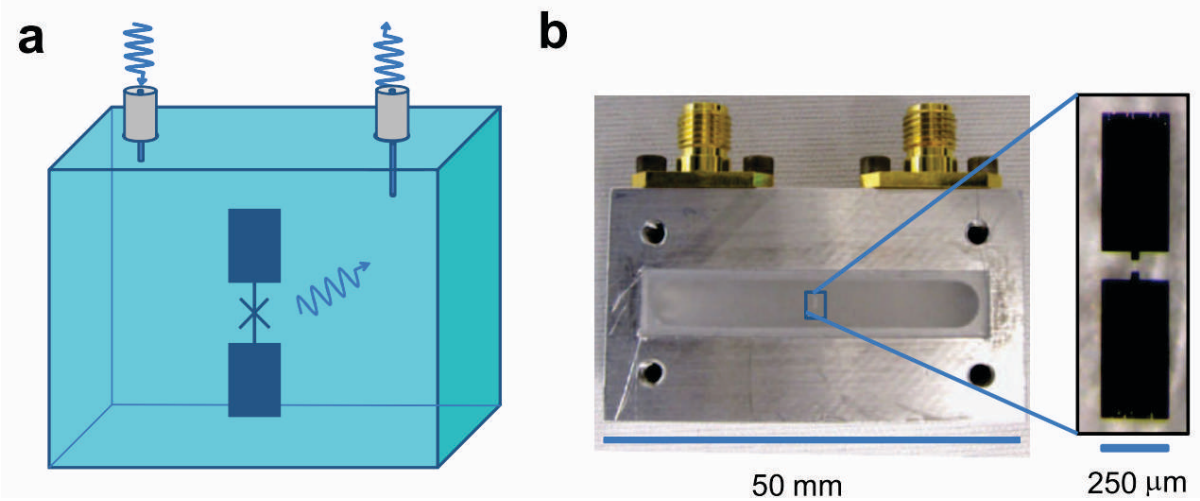


Figure 3.1: **The three-dimensional transmon.** (a) Cartoon of a qubit in a three-dimensional cavity. (b) Transmon qubit inside an Al microwave cavity. The qubit acts as a dipole antenna, allowing qubit-cavity coupling. This figure reproduced from Ref. [**?**]

enabled a new class of physics experiments. In our group, as our primary goal was to research a surface code quantum computer, we sought to bring the coherence of the three-dimensional transmon into a frequency-tunable transmon in a planar geometry. We believed that planar integrated circuits would be critical to realizing a complex quantum computing architecture, as sophisticated control circuitry was likely necessary. This research direction seemed possible as recent results demonstrated highly coherent planar resonators with quality factors up to 1.6 million, implying a T_1 of 50 μs at 5 GHz [**?**], where improvements in materials greatly reduced capacitive losses, and junctions were known to be intrinsically low loss. We believed building such a qubit as possible, and would be very competitive with modern qubit technology.

The scope of this work goes beyond integrating low-loss materials into known transmon geometries, as traditional cQED architectures did not seem directly compatible with the surface code. A device displaying the cannonical cQED architecture is shown in Fig. 3.2. This device

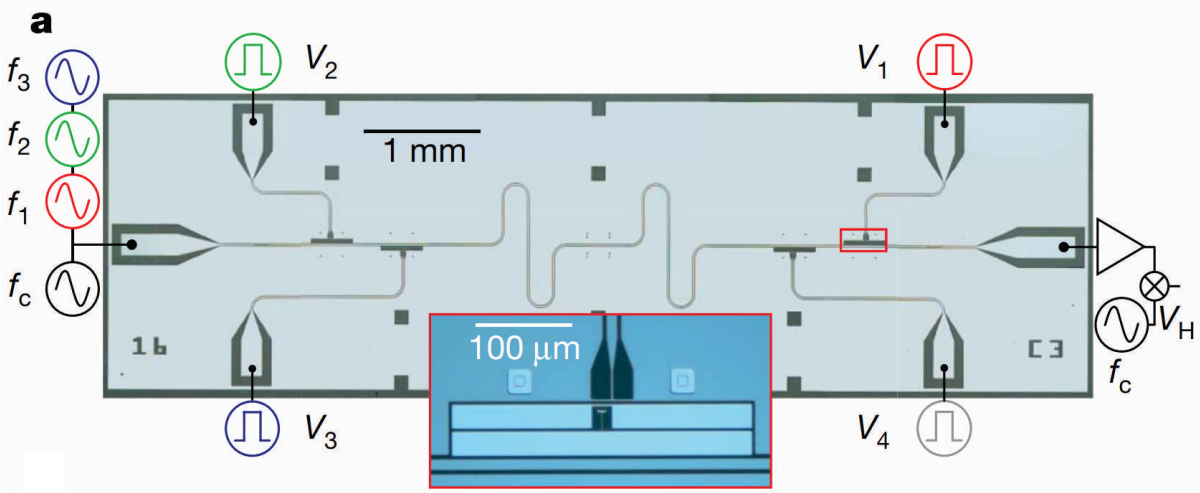


Figure 3.2: **The cannonical cQED architecture.** This figure reproduced from Ref. [**?**]. Four transmon qubits coupled to a central quantum bus. The quantum bus provides qubit-qubit coupling, qubit readout (f_c), and routes microwaves for qubit manipulations (f_1, f_2, f_3).

consists of four transmon qubits coupled to a quantum bus, and can be thought of as four atoms in a cavity. The central feature of this device is the quantum bus, which has simultaneously has three separate functions. First, this device is used for qubit readout. As each qubit is coupled to the bus, the frequency of the bus resonator depends on qubit state, and can be probed with external microwaves. Second, qubit-qubit coupling is realized through virtual excitations of the bus. Third, the bus brings in external microwaves for qubit manipulations.

This cQED architecture, though beautiful in its simplicity and demonstratively powerful for a variety of physics experiments [**?** **?** **?**], is problematic for the surface code. First, as all qubits are coupled to the bus, the readout scheme is joint: qubits cannot be individually measured. The surface code requires measurements of measure qubits independent of data qubits. Second, the surface code requires nearest neighbor connectivity. This architecture has all qubits simultaneously coupled through the bus, a higher level of connectivity than required,

and thus prone to unwanted interactions. Additionally, it is not immediately obvious how to extend this architecture past a few qubits. Thus, we began investigation the Xmon design. It is worth noting that other research groups recently have used a modified cQED scheme to implement surface code achitectures [**?** **?**].

3.3 Design

Our device is shown in Fig. 3.3a, formed by patterning the Al metal (light), exposing the sapphire substrate (dark). The qubit is the cross-shaped device. We design the qubit with high-quality coplanar waveguide capacitors, motivated by the recent advances with superconducting resonators, yielding a modular design with straightforward connectivity. Its four arms connect to separate elements, each having a different function: a coplanar waveguide resonator for readout on the top, a quantum bus resonator on the right to mediate coupling to other qubits, XY control on the left to excite the qubit state, and Z control on the bottom to tune the qubit frequency. The cross is the qubit capacitor, which connects at the bottom to the tunable Josephson junction, formed by the rectangular ring-shaped superconducting quantum interference device (SQUID), see Fig. 3.3b. The rectangular ring is intersected by two identically sized Al tunnel junctions (blue in Fig. 3.3b). The electrical circuit is equivalent to that of a grounded transmon [**?**], with the capacitor in parallel with the tunable junction (Fig. 3.3c). In a clear departure from the traditional floating transmon with an interdigitated capacitor (IDC) [**?**], we chose to form the qubit capacitor by intersecting two coplanar waveguide lines.

In prior work, we showed that highly coherent coplanar waveguide resonators can be fab-

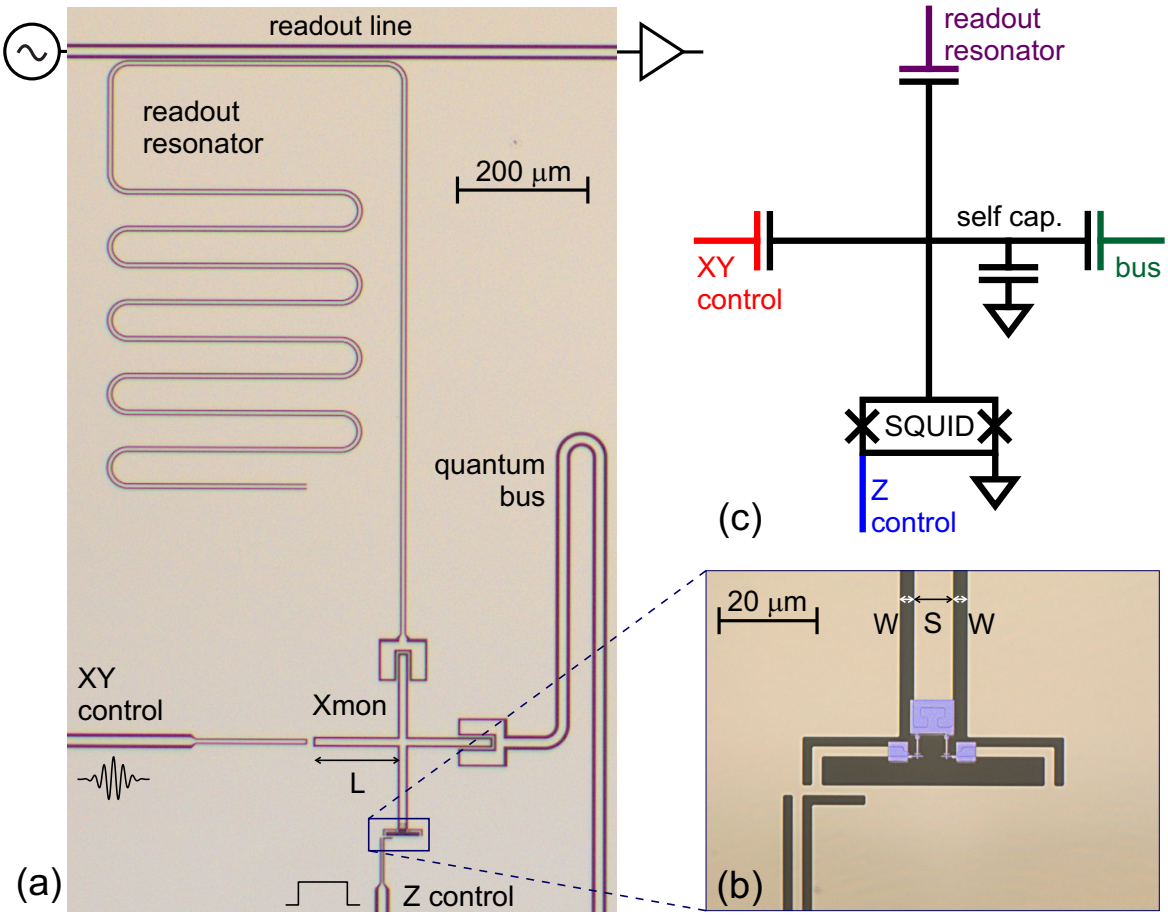


Figure 3.3: (Color online) (a) Optical micrograph of the planar 'Xmon' qubit, formed by the Al superconducting film (light) and the exposed sapphire substrate (dark). The qubit is capacitively coupled to a quarter wave readout resonator (top), a quantum bus resonator (right), a XY control line (left), and inductively coupled to a Z control line (bottom). The Xmon arm length is L. (b) The inset shows the shadow evaporated Al junction layer in false color (blue). The junction size is 0.30×0.20 μm^2. The capacitor central linewidth is S, the gap width is W. (c) The electrical circuit of the qubit.

ricated, having quality factors of about 1.5×10^6 at the single photon occupation level. These resonators were made from molecular beam epitaxy (MBE) Al on oxygen-cleaned sapphire [**?**]. This shows that a straightforward path to high coherence comes from a combination of I) MBE Al as high quality material, II) coplanar waveguides having low radiative loss, and III) embedding in a groundplane. We therefore embed the qubit in an uninterrupted groundplane, with thin Al lines at the capacitor ends tying the groundplanes together; this suppresses parasitic slotline modes in the control lines and resonators as well.

Connectivity is accomplished by coupling each of the qubit's arms to a distinct element with specific functionality. Three of the connections are easily made with a coupling capacitor, as the qubit is connected to ground. An advantage of this approach is that each coupling can be individually tuned and optimized. To this end, we have also separated out qubit control. The XY control drive line is connected with a coupling $C_c = 60$ aF, which allows us to excite the qubit state in 10 ns but hardly affects coherence, with an estimated T_1 of 0.3 ms. The Z control also combines speed and coherence. The drive line is galvanically connected to the SQUID to allow for a large inductive coupling with a mutual inductance of $M = 2.2$ pH. We are able to rapidly detune the qubit on the timescale of a nanosecond (Appendix A). The measured parasitic coupling between the Z line and the qubit gives an estimated T_1 of $\sim$30 ms [**?**].

We believe that the large increase in the qubit coherence relies critically on the *combination* of changes to the qubit design; implementing just one or two of these changes in isolation would not yield a significant improvement. With this experimental nature in mind, we name our qubit the 'Xmon'. While the cross-shaped qubit capacitor may emphasize this name, more

arms can be added to allow for more connectivity.

3.4 Coherence

We find a dramatic increase in Xmon energy coherence compared to the traditional planar transmon, measuring decay times up to $T_1 = 44$ μs, see Fig. 3.4a. We find Ramsey and spin echo phase coherence times up to $T_2^* = 15$ μs and $T_2 = 20$ μs at the flux insensitive point, respectively, see Fig. 3.4b; $T_1 = 18$ μs at this point. The dephasing envelopes follow an exponential decay, measured using tomography. The first pulse is $X_{\pi/2}$, followed by $X_{\pi/2}$, $X_{-\pi/2}$, $Y_{\pi/2}$, or $Y_{-\pi/2}$ (see inset), producing fringes with different phases. The limit of $T_2 = 2T_1$ has not been reached [**?**], indicating additional dephasing. This, as well as dephasing away from the flux insensitive point, is presently under investigation.

The qubits had ground to excited state transition frequencies around 6 GHz when unbiased, nonlinearities around 230 MHz, and a ratio of Josephson to charging energy $E_J/E_C \sim 95$. In initial experiments, we employed a dispersive, high-power single-shot readout scheme with a 70-85% fidelity [**?**]. The readout resonator frequencies used were 6.4-6.7 GHz, with loaded quality factors of $Q_l = 10^4$, and resonator-qubit coupling strengths of approximately 40 MHz. Measurements were done in a dilution refrigerator with a base temperature of 30 mK, with multistage infrared shielding [**?**]. Magnetic fields were reduced by room temperature and cryogenic magnetic shields, with nonmagnetic microwave connectors [**?**].

The results in Fig. 3.4 show that tunable superconducting qubits with a planar geometry can have T_1 values in excess of 40 μs. In fact, this T_1 corresponds to the MBE Al resonator quality

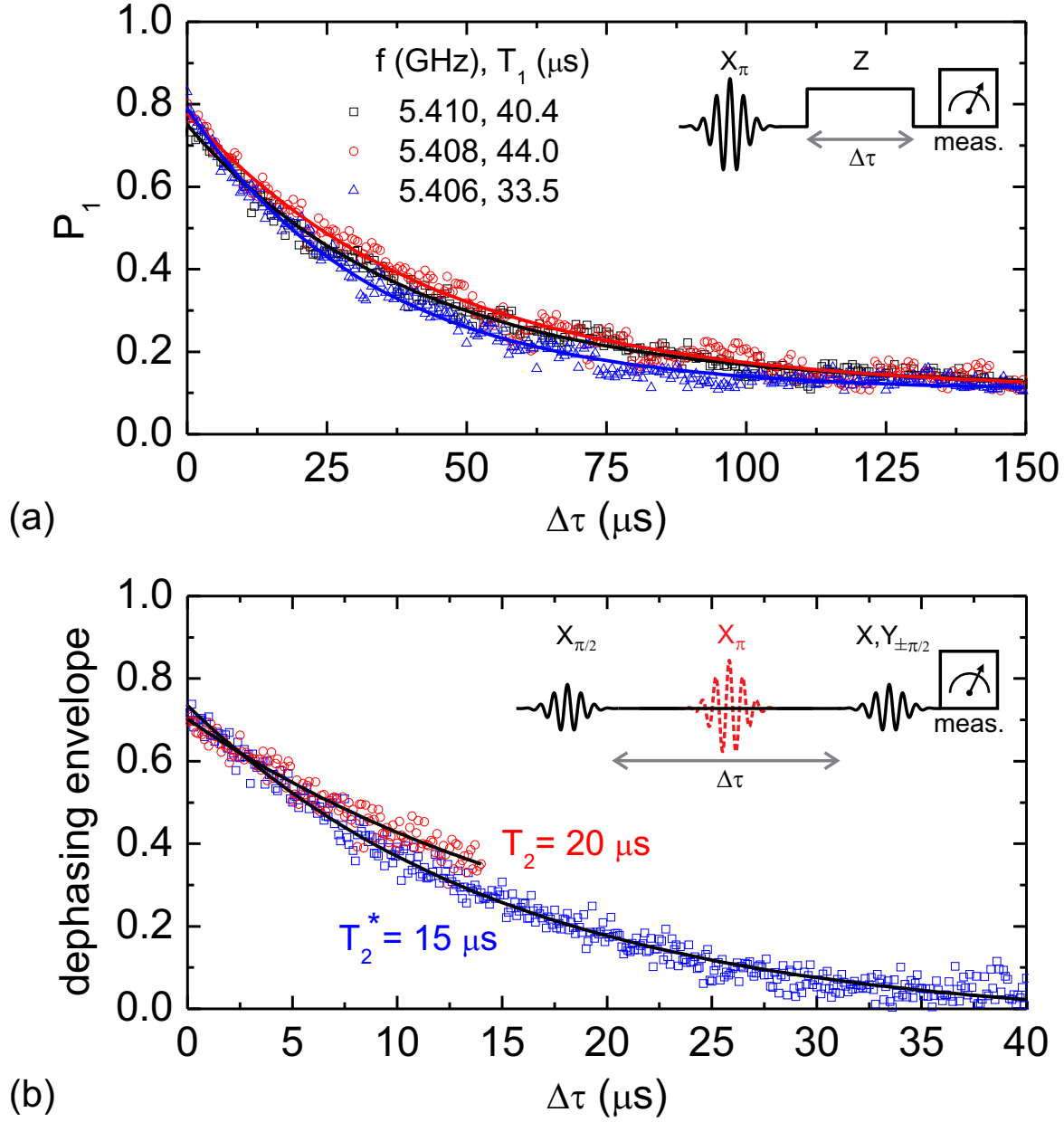


Figure 3.4: (Color online) (a) Qubit energy decay at three nearby frequencies ($S = W = 16$ μm [**?**]). The qubit frequency is adjusted by applying a rectangular pulse with length $\Delta\tau$ on the Z line. The pulse sequence is shown in the inset. (b) Ramsey T_2^* and spin echo T_2 dephasing envelopes at the flux insensitive point ($T_1 = 18$ μs), measured by phase tomography ($S = W = 24$ μm). The inset shows the pulse sequence; for the spin echo we apply a refocussing pulse (dashed). We apply four phases to the last pulse for phase tomography, to measure the decay envelope. Spin echo measurements are limited in time by electronics to 14 μs.

factors [**?**], for which $T_1 = Q/\omega$ is also about 40 μs (however, the average performance is slightly less). The combination of long energy and phase coherence times compares well with previously reported values for planar superconducting Al qubits: for transmons $T_1 = 9.7$ μs and $T_2^* = 10$ μs [**?**] and recently 20-30 μs [**?**], for charge qubits $T_1 = 200$ μs and $T_2^* = 0.07$ μs [**?**], for flux qubits $T_1 = 12$ μs and $T_2^* = 2.5$ μs [**?**], and for the fluxonium $T_1 = 10$ μs and $T_2^* = 2$ μs [**?**]. In fact, the Xmon approaches the long coherence found in 3D transmons [**? ?**]. Very recently, TiN planar devices have shown long coherence [**? ?**], encouraging using Xmon geometries with this material.

3.5 Device design and coherence history

Figure 3.5 shows the history of transmon and Xmon devices from January 2012 to July 2013, with best their coherence times and SQUID loop designs. Improvements in materials, microwave engineering, and simplifications in fabrication saw improvements in energy relaxation times from 1.6 μs to 42 μs in roughly one year, and up to 57 μs soon after in a five qubit device.

3.6 Incoherent defects: Data

We find that the energy relaxation depends on qubit frequency. As shown in Fig. 3.4a, we find T_1 values from 34 to 44 μs in a 4 MHz band near 5.4 GHz. In order to elucidate this further, we performed a spectroscopic scan on the qubit, shown in Fig. 3.6a. The qubit frequency displays the expected dependence on applied flux Φ [**?**], varying smoothly without visible

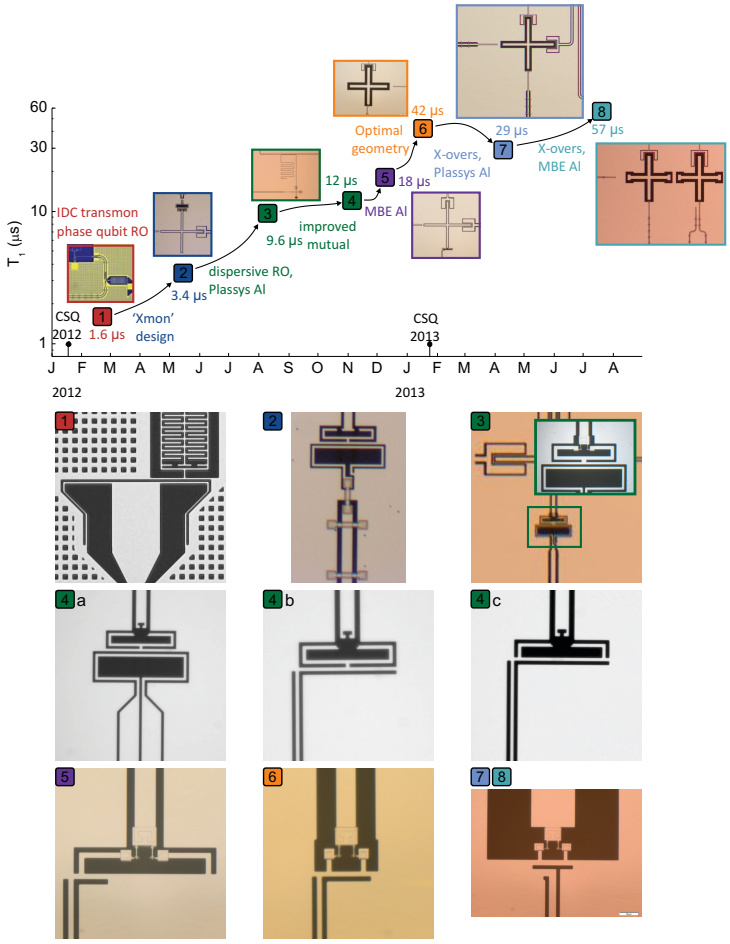


Figure 3.5: **Xmon coherence and design** $vs.$ **time.** (TOP)(1) Initial transmon design, transmon with phase qubit readout. Limited by overcoupling to control lines. (2) Initial Xmon design, limited by poor materials due to sputter deposited Al and complex fabrication. (3) Two layer Xmon design. Incorporated high quality e-beam evaporated Al, removed dirty processing steps, and uses wirebond crossovers. (4) Introduction of galvanic bias line, improve control. (5) Introduction of MBE Al. (6) Optimized geometry and MBE Al. (7) Introduction of SiO_2 crossovers to reduce stray microwave modes by improving ground plane connectivity. (7,8) Five qubit device incorporating all design and materials improvements. (BOTTOM) (1) Initial transmon design. Bias line has large footprint and poor coupling (<0.2 pH). (2) Initial Xmon SQUID loop design. Modest coupling of 0.4 pH, but reliant on crossovers to balance current. (3) Single layer version of previous design, but lack of crossovers reduced coupling (<0.2 pH). (4a-c) Design array of bias lines to find optimal geometry. (4c) Strong coupling (7 pH) achieved by galvanically connecting bias line to SQUID loop, while maintaining ground plane connectivity. (5) Reduced coupling to 2 pH, reducing sensitivity to noise on bias line by an order of magnitude. (6) 2 pH coupling with smaller SQUID loop, however larger transmon geometries have microwave coupling to bias line due to asymmetry. (7,8) Optimal and symmetric 1.7 pH design with greatly reduced stray coupling.

splittings, indicating that strongly coupled defects, which manifest as avoided level crossings [**?**], are virtually absent. We then performed a quantum analogue of time-resolved spectroscopy (swap spectroscopy [**?**]), shown in Fig. 3.6b. The pulse sequence is shown in Fig. 3.4a. The probability of the excited state (color) is plotted for $\Delta\tau$ from 100 ns to 150 μs (logarithmic vertical scale) and qubit frequencies from 4 to 6 GHz. We find that the probability decays exponentially, but with a fine structure of variable energy relaxation, and distinct peaks in the energy decay rate (Fig. 3.6c). We do not observe any "chevron" interference patterns [**?**], where the quantum state coherently swaps back and forth, implying no defects interact coherently with the qubit. After cycling the temperature to 4.2 K the fine structure is altered, but the overall image remains unchanged. We count approximately 30 regions with reduced coherence ($T_1 < 8$ μs) per GHz in Fig. 3.6b.

We explored the dependence of the qubit coherence time on capacitor geometry, using six different designs; the central line width S, gap width W and arm length L were varied, while the capacitance value [**?**] and junction parameters are kept the same. The parameters are listed in Table 3.1; see Appendix A for a micrograph. We find that the swap spectroscopy measurements of the different designs share the same characteristics as shown in Fig. 3.6b: a fine structure with varying exponential decay. The energy relaxation times extracted from the

Table 3.1: Geometric parameters for the Xmon qubit capacitors as defined in Fig. 3.3 along with their frequencies. Groups of three qubits indicate that the devices are on the same chips.

S (μm)	8	16	24	8	16	24
W (μm)	4	8	12	8	16	24
L (μm)	130	130	130	165	165	165
$f_{10,\max}$ (GHz)	6.094	6.158	6.071	6.080	5.883	5.846
nonlinearity (MHz)	224	228	222	220	225	223

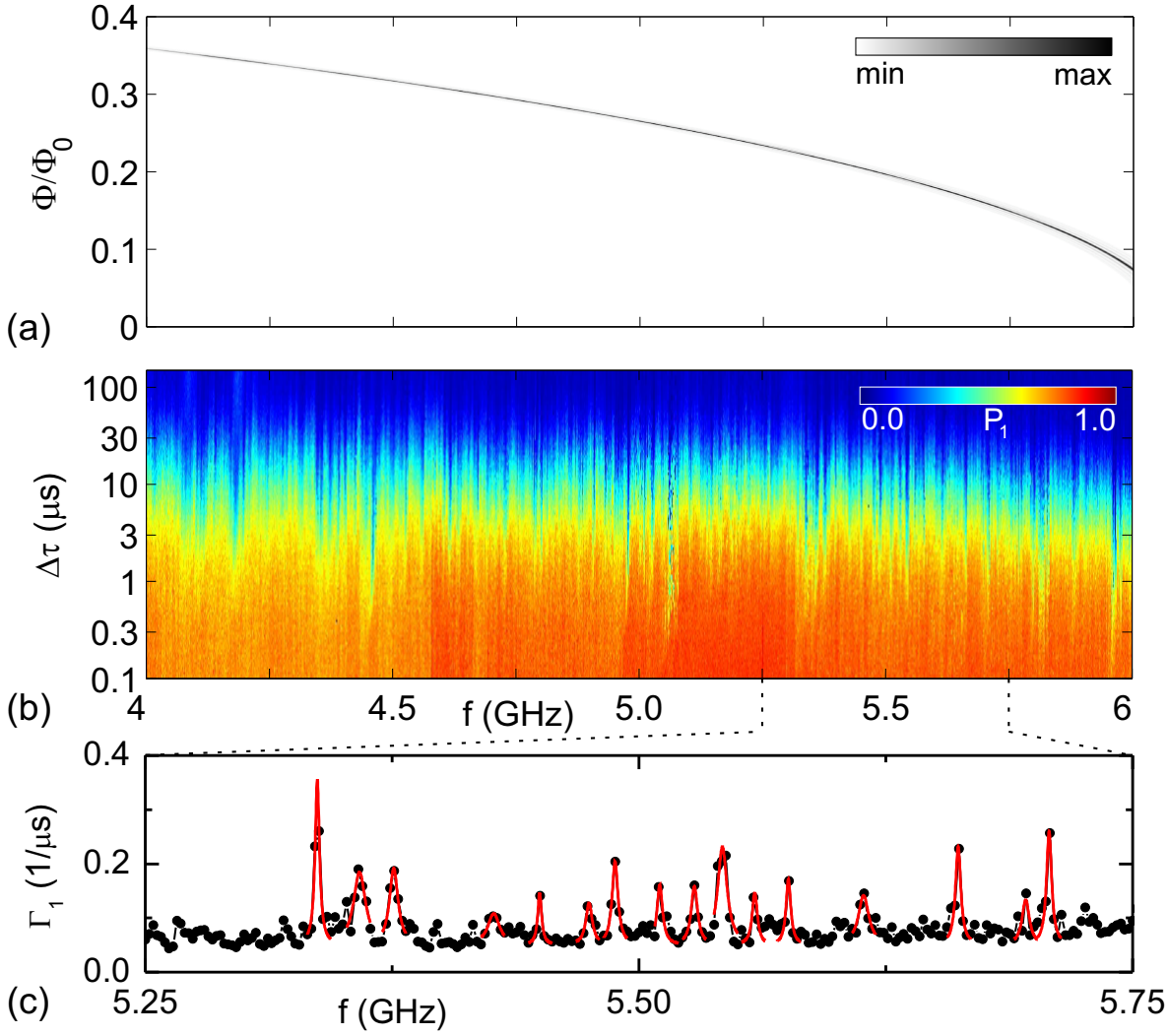


Figure 3.6: (Color online) (a) Qubit spectroscopy for device with S,W=8,8 μm. A smooth curve is formed by the high transmission (grey), measured on resonance with the readout resonator, which indicates when the qubit is excited. (b) Swap spectroscopy of the same qubit. The qubit is detuned from 4 to 6 GHz (stepsize 2 MHz), and the delay time is varied from 100 ns to 150 μs. See the inset of Fig. 3.4a for the pulse sequence. (c) Qubit relaxation rates, extracted from the data in (b). The peaks are fitted to Eq. 3.1 (solid lines).

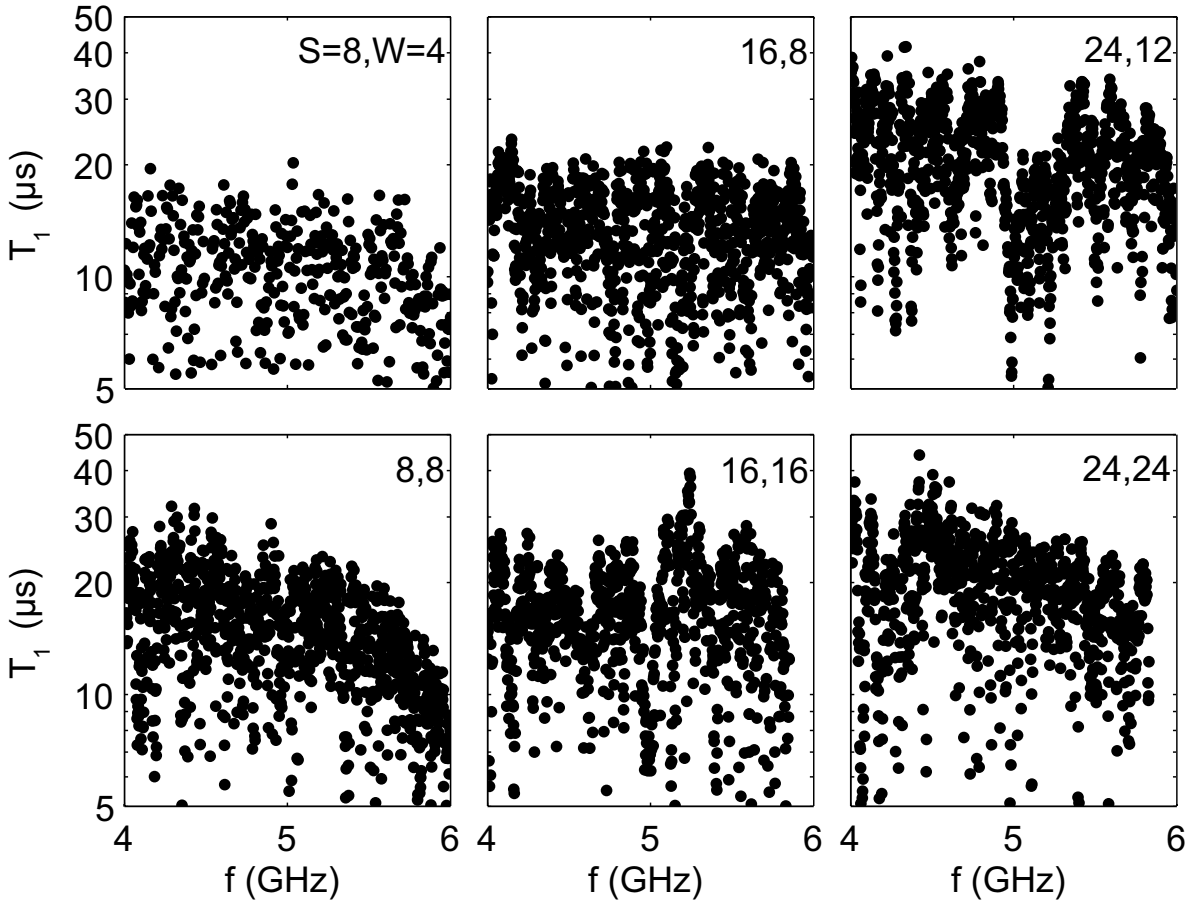


Figure 3.7: Frequency dependence of T_1 for six qubits with different S and W (see Table 3.1). The frequency stepsize is 5 MHz for S,W = 8,4 μm and 2 MHz otherwise. See Appendix A for the corresponding decay rates.

measurements are shown in Fig. 3.7. The overall energy relaxation time increases with width: when changing S,W from 8,4 μm to 16,8 μm and 24,12 μm, the T_1 improves from a band of values between 8-15 to 10-20, and 20-40 μs, respectively. Importantly, both the upper and lower bounds on T_1 increase. This is repeated in the qubits with S,W ranging from 8,8 μm to 16,16 μm and 24,24 μm. The reduction of T_1 at frequencies approaching 6 GHz is due to Purcell decay into the readout resonator [**?**]. We emphasize that these T_1 values are obtained in multiqubit chips with control wiring.

The improvement of T_1 with increasing width is consistent with previous experiments on superconducting resonators [**?** **?**]. Loss arises from the electric fields coupling to two-level systems with dipole moments [**?**], which reside predominantly in surface oxides and interfaces. This loss depends on the participation ratio, which depends on the electric field distribution [**?**]. Widening the capacitor reduces the surface participation, a natural explanation for the approximately linear increase in average T_1 with width in Fig. 3.7. On the other hand, the peaks in the decay rate are reminiscent of experiments with phase qubits [**?**], where localized features in the frequency dependence occur when the qubit couples strongly to two-level defects, often giving rise to splitting of the qubit frequency and the chevron-shaped signature of coherent swapping. However, the exponential decay in the Xmon qubit, with no signatures of swapping or splitting, suggests a different energy relaxation mechanism.

3.7 Incoherent defects: Theory and numerics

Here, we show how surface defects near the metal edges of the capacitor provide a natural explanation for the peaks in the energy decay. As indicated by the data, the key point is that loss arises from the qubit interacting with a sparse bath of incoherent, weakly-coupled defects, giving rise to incoherent decay [**?**]. The sharp frequency dependence as well as the changes in fine structure when cycling to 4.2 K are consistent with defects. The absence of chevrons and qubit frequency splittings corresponds to incoherent interaction. The lower and upper bounds of T_1 increasing with capacitor dimension indicates the defects reside in the capacitor.

We model a quantum system consisting of a qubit, with a frequency-independent loss rate $\Gamma_{1,\mathrm{Q}}$ and pure dephasing rate $\Gamma_{\phi,Q}$, and a two-level defect with decoherence rate $\Gamma_{1,\mathrm{D}}$ and dephasing rate $\Gamma_{\phi,\mathrm{D}}$ (Appendix A); we assume Markovian decoherence. When $\Gamma_{1,\mathrm{D}}$ exceeds the coupling strength g, coherent swapping vanishes and an incoherent, exponential decay appears. From a two-spin Hamiltonian (Appendix A) we derive the qubit energy relaxation rate Γ_1 (in the limit $\Gamma_{1,\mathrm{D}} > g > \Gamma_{1,\mathrm{Q}}$)

$$\Gamma_1 = \frac{2g^2\Gamma}{\Gamma^2 + \Delta^2} + \Gamma_{1,\mathrm{Q}}, \tag{3.1}$$

with detuning Δ, and $\Gamma = \Gamma_{1,\mathrm{D}}/2 + \Gamma_{\phi,\mathrm{D}} + \Gamma_{1,\mathrm{Q}}/2 + \Gamma_{\phi,\mathrm{Q}}$. Hence, each uncorrelated defect adds a single Lorentzian to the energy decay rate. We can roughly estimate g for a surface defect with dipole moment $p \sim 1$ D [**?**] at a distance x away from the metal edge. With the electric field given by $E = B/\sqrt{x}$ [**?**], and a parameter B from numerical simulations [**?**], we arrive at

$g/2\pi \sim 0.1$ MHz ($g = pE$) for $x = 3$ nm. We apply our model to Fig. 3.6c. The peaks in decay rate can be described by a set of Lorentzians, with $1/(\Gamma_{1,\mathrm{D}}/2 + \Gamma_{\phi,\mathrm{D}}) \sim 50 - 100$ ns, consistent with defect decay rates measured in similar systems [**?** **?**], and with $g/2\pi \gtrsim 0.2$ MHz, agreeing with incoherent loss.

We can also estimate the number of individually resolvable defects using two-level system physics developed for junctions. The substrate-metal interface in our devices was thoroughly cleaned [**?**], hence we assume that the bulk of strongly-coupled defects resides in the metal- and substrate-air interfaces, as they have the highest participation ratios [**?**]. The defect density for AlO_x in tunnel barriers has been established in measurements with phase qubits [**?**], with the distribution over dipole moment given by $\rho_0 \sqrt{1 - p^2/p_{\mathrm{max}}^2}/p$, with $\rho_0 \approx 10^2/\mu\mathrm{m}^3/\mathrm{GHz}$, and the maximum dipole moment $p_{\mathrm{max}} = 6$ D. We take these numbers as representative and assume a 3 nm thick dielectric layer with defects [**?**]. The number of defects with coupling strength greater than g_{min} is then given by

$$N = \iint \rho_0 \frac{\sqrt{1 - p^2/p_{\mathrm{max}}^2}}{p} \Theta \left[p|E(\vec{r})| - g_{\mathrm{min}} \right] dp d\vec{r}, \tag{3.2}$$

with Θ the unit step function and $E(\vec{r})$ the electric field at position $\vec{r}$. Simulations using Eq. 3.2 as well as Monte Carlo simulations indicate $N \sim 30 - 50$/GHz, for $g_{\mathrm{min}}/2\pi \sim 0.2$ MHz, see Appendix A. We emphasize that the simulations connect g_{min} to N with values which are close to what is observed experimentally. The simulations also indicate that the bulk of strongly coupled defects reside within a $\sim$100 nm distance from the etched edges, where the electric fields are largest. In addition, the simulated qubit decay rate reproduces the experimentally

observed features, showing both the peaks and background variation.

The good quantitative comparison between model and experiment gives compelling evidence that a sparse bath of incoherent defects plays a major role in loss in highly coherent qubits. Our results may also explain previously reported anomalous behavior in planar transmon qubits with long coherence, for which the T_1 has been reported to vary significantly between qubits, even on the same chip [**? ?**]. This is consistent with a sparse bath of incoherent defects limiting the coherence, as in Fig. 3.6.

3.8 Conclusion

In conclusion, we demonstrate energy coherence times exceeding 40 μs in tunable, planar superconducting qubits. We have achieved this using a geometry with low radiative dissipation and high quality materials. At these high coherence levels we identify a novel decoherence mechanism, loss from a sparse bath of incoherent defects, which is apparent in the swap spectroscopy. Our qubits combine long coherence, easy interconnectivity, and fast control, providing a key ingredient for the implementation of an on-chip surface code quantum computer.

Chapter 4

Superconducting quantum circuits at the surface code threshold for fault tolerance[1]

A quantum computer can solve hard problems - such as prime factoring [**?** **?**], database searching [**?** **?**], and quantum simulation [**?**] - at the cost of needing to protect fragile quantum states from error. Quantum error correction [**?**] provides this protection, by distributing a logical state among many physical qubits via quantum entanglement. Superconductivity is an appealing platform, as it allows for constructing large quantum circuits, and is compatible with microfabrication. For superconducting qubits the surface code [**?**] is a natural choice for error correction, as it uses only nearest-neighbour coupling and rapidly-cycled entangling gates. The gate fidelity requirements are modest: The per-step fidelity threshold is only about 99%. Here, we demonstrate a universal set of logic gates in a superconducting multi-qubit processor,

[1]The following was published as: "Superconducting quantum circuits at the surface code threshold for fault tolerance", Rami Barends, Julian Kelly, et al. *Nature* 508, 500-503 (2014).

achieving an average single-qubit gate fidelity of 99.92% and a two-qubit gate fidelity up to 99.4%. This places Josephson quantum computing at the fault-tolerant threshold for surface code error correction. Our quantum processor is a first step towards the surface code, using five qubits arranged in a linear array with nearest-neighbour coupling. As a further demonstration, we construct a five-qubit Greenberger-Horne-Zeilinger (GHZ) state [**?** **?**] using the complete circuit and full set of gates. The results demonstrate that Josephson quantum computing is a high-fidelity technology, with a clear path to scaling up to large-scale, fault-tolerant quantum circuits.

Fault-tolerance in the surface code is achieved by placing physical qubits in a checkerboard pattern, with white squares representing data and black squares representing measurement qubits that detect errors. To perform this detection, each measurement qubit needs to interact with its four neighbouring data qubits. All that are needed for these interactions are single- and two-qubit gates, with sufficiently high fidelity. The high fidelity demonstrated here is achieved through a combination of coherent qubits, a straightforward interconnection architecture, and a novel implementation of the two-qubit controlled-phase (CZ) entangling gate. The CZ gate uses a fast but adiabatic frequency tuning of the qubits [**?**] that mimises error [**?**].

Here, the tuneable nature of the qubits and their entangling gates provides, remarkably, both high fidelity and fast control. Previous demonstrations of two-qubit gates achieving better than 99% fidelity used fixed-frequency qubits: Systems based on nuclear magnetic resonance and ion traps have shown two-qubit gates with fidelities of 99.5% [**?**] and 99.3% [**?**]. Recently,

for a five qubit ion trap [**?**] and three qubit superconducting system [**?**], two-qubit entangling gate fidelities of 95% and 96% were reported.

4.1 Device

Superconductivity allows for building large quantum integrated circuits as the electrons are condensed into a single macroscopic quantum state. As shown in Fig. 4.1, we have designed a processor for pioneering tests with the surface code; it consists of five Xmon transmon qubits with nearest-neighbour coupling, arranged in a linear array. The cross-shaped qubit [**?**] offers a nodal approach to connectivity while maintaining a high level of coherence (see Appendix B for decoherence times). Here, the four legs of the cross allow for a natural segmentation of the design into coupling, control and readout. We chose a modest inter-qubit capacitive coupling strength of $g/2\pi = 30$ MHz and use alternating qubit idle frequencies of 5.5 and 4.7 GHz, enabling a CZ gate in 40 ns when two qubits are brought near resonance, while minimising the effective coupling to 0.3 MHz when the qubits are at their idle points. Rotations around the X and Y axes in the Bloch sphere representation are performed using pulses on the microwave (XY) line, while Z axis rotations are achieved by a flux-bias current on the frequency-control (Z) line. We use dispersive measurement [**?**] where each qubit is coupled to a readout resonator, each with a different resonance frequency, enabling simultaneous readout using frequency-domain multiplexing through a single coplanar waveguide [**?**]. The modularity of this architecture makes it straightforward to integrate more qubits in the circuit.

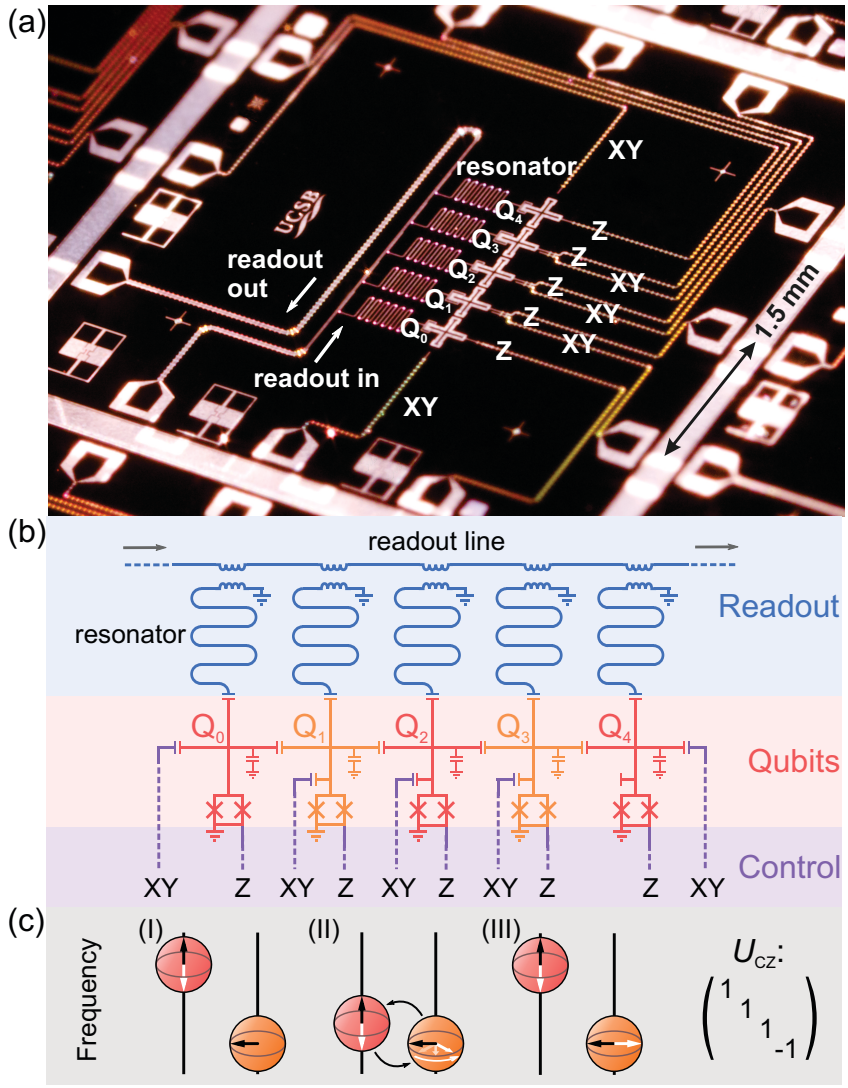


Figure 4.1: **Architecture.** (a) Optical image of the integrated Josephson quantum processor, consisting of Al (dark) on sapphire (light). The five cross-shaped devices are the Xmon variant of the transmon qubit [**?**], labelled $Q_0 - Q_4$, placed in a linear array. To the left of the qubits are five meandering coplanar waveguide resonators used for individual state readout. Control wiring is brought in from the contact pads at the edge of the chip, ending at the right of the qubits. (b) Circuit diagram. Our architecture employs direct, nearest-neighbour coupling of the qubits (red/orange), made possible by the nodal connectivity of the Xmon qubit. Using a single readout line, each qubit can be measured using frequency-domain multiplexing (blue). Individual qubits are driven through capacitively-coupled microwave control lines (XY), and frequency control is achieved through inductively-coupled dc lines (Z) (violet). (c) Schematic representation of an entangling operation using a controlled-Z gate with unitary representation U_{CZ}: (I) Qubits at rest, at distinct frequencies with minimal interaction. (II) When brought near resonance, the state-dependent frequency shift brings about a rotation conditional on the qubit states. (III) Qubits are returned to their rest frequency.

4.2 Randomized Benchmarking

We characterise our gate fidelities using Clifford-based randomised benchmarking [**?** **?** **?**]. The Clifford group is a set of rotations that evenly samples the Hilbert space, thus the benchmarking averages across errors. For the single-qubit case the Cliffords are comprised of π, $\pi/2$ and $2\pi/3$ rotations, see Appendix B. In randomised benchmarking, a logic gate is characterised by measuring its performance when interleaved with many random sequences of gates, making the measured fidelity resilient to state preparation and measurement (SPAM) errors. We first perform a control experiment on a ground-state qubit by: I) generating a random sequence of m Cliffords, II) appending the unique recovery Clifford (C_{r}) that inverts the sequence, and III) averaging the experimental sequence fidelity, the final ground state population, over k different sequences [**?** **?**]. The resulting reference sequence fidelity F_{ref} is fit to $F_{\mathrm{ref}} = A{p_{\mathrm{ref}}}^m + B$, where p_{ref} is the sequence decay, and state preparation and measurement errors are captured in the parameters A and B. The average error per Clifford of the reference is given by $r_{\mathrm{ref}} = (1 - p_{\mathrm{ref}})(d-1)/d$, with $d = 2^{N_{\mathrm{qubits}}}$. We then measure the fidelity of a specific gate by interleaving it with m random Cliffords. The sequence decay p_{gate} gives the gate error $r_{\mathrm{gate}} = (1 - p_{\mathrm{gate}}/p_{\mathrm{ref}})(d-1)/d$.

4.3 Single qubit gates

The benchmarking results for the single-qubit gates are shown in Fig. 4.2. We generate the Cliffords using microwave pulses, from a basis set of π and $\pi/2$ rotations around the X and

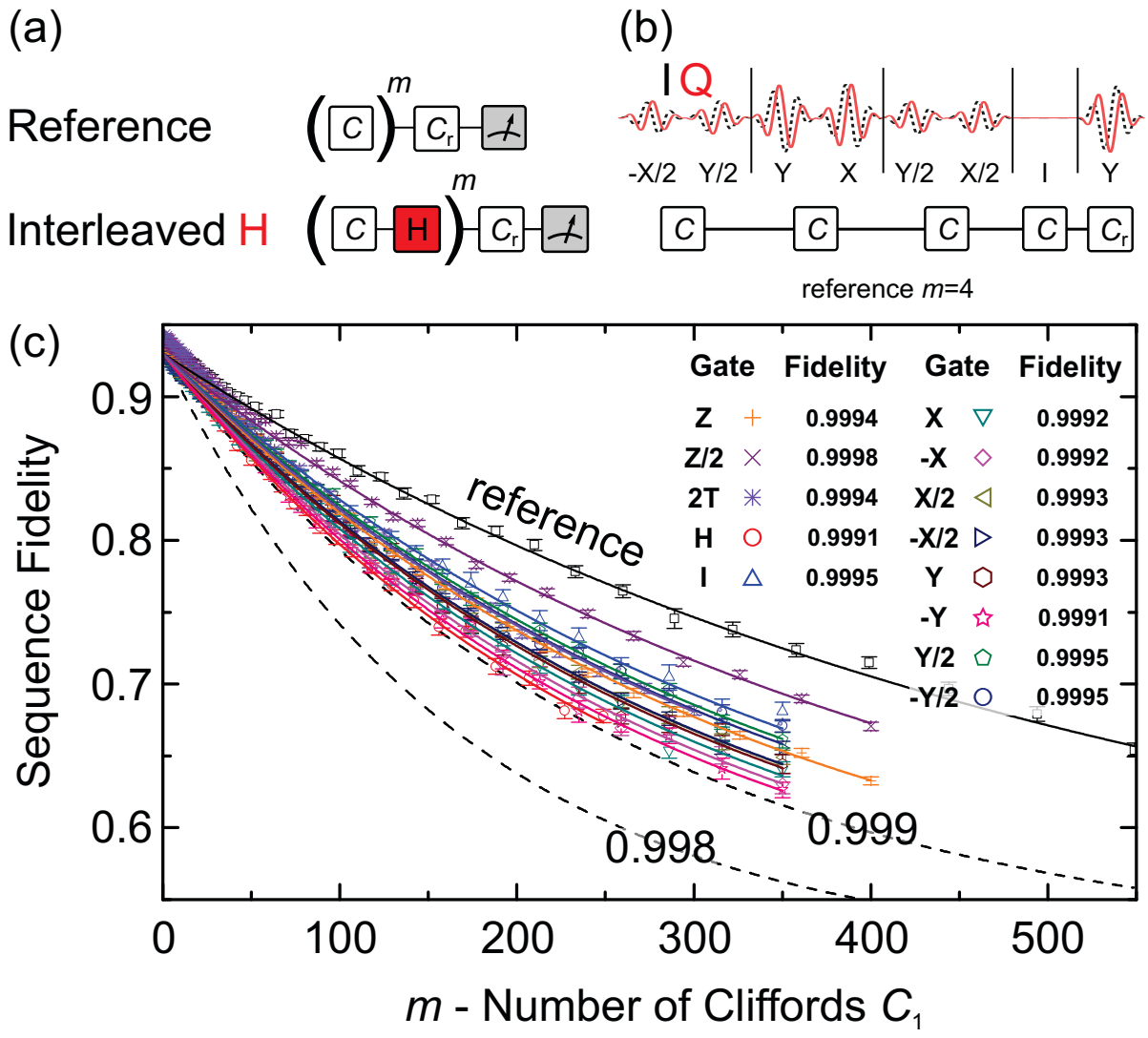


Figure 4.2: **Single qubit randomised benchmarking.** (a) A reference experiment is performed by generating a sequence of m random Cliffords, which are inverted by the recovery Clifford C_r. A specific gate (H) is tested using a sequence that interleaves H with m random Cliffords. The difference between interleaved and reference decay gives the gate fidelity. (b) Representative pulse sequence for a set of four Cliffords and their recovery, generated with π and $\pi/2$ rotations about X and Y, displaying both the real (I) and imaginary (Q) microwave pulse envelopes before up-conversion by quadrature mixing to the qubit frequency. (c) Randomised benchmarking measurement for the set of single-qubit gates for qubit Q_2, plotting reference and gate fidelities as a function of the sequence length m; the fidelity for each value of m was measured for $k = 40$ different sequences. The fit to the reference set yields an average error per Clifford of $r_{\text{ref}} = 0.0011$, consistent with an average gate fidelity of $1 - r_{\text{ref}}/1.875 = 0.9994$ (Appendix B). The dashed lines indicate the thresholds for exceeding gate fidelities of 0.998 and 0.999. The fidelities for each of the single-qubit gates are tabulated in the legend, we find that all gates have fidelities greater than 0.999. Standard deviations are typically $5 \cdot 10^{-5}$.

Y axes (Appendix B). We benchmark X and Y axis π and $\pi/2$ rotations, the Hadamard gate (implemented with Y/2 followed by X), and Z axis rotations using pulses on the frequency control line. From the data in Fig. 4.2 we extract the individual gate fidelities listed in the legend. We find an average fidelity of 99.92 % over all gates and qubits (Appendix B). The best fidelities are achieved by optimising the pulse amplitude and frequency, and minimising 2-state leakage [**? ?**].

We have also measured the performance when simultaneously operating nearest or next-nearest qubits [**?**], the qubits are at dissimilar idle frequencies to minimise coupling. The fidelities are essentially unchanged, with small added errors $< 2 \cdot 10^{-4}$ (Appendix B), showing a high degree of addressability for this architecture.

4.4 Controlled-Phase gate

The two-qubit CZ gate is implemented by tuning one qubit in frequency along a "fast adiabatic" trajectory which takes the two-qubit $|11\rangle$ state close to the avoided-level crossing with the $|02\rangle$ state, yielding a state-dependent relative phase shift. This implementation is the natural choice for weakly anharmonic, frequency-tunable qubits, as the other computational states are left unchanged [**? ? ?**]. Having the CZ gate adiabatic as well as fast is advantageous. An adiabatic trajectory is easily optimised and allows for exponentially suppressing leakage into the non-computational $|02\rangle$-state with gate duration. Having a fast CZ gate minimises the accumulation of errors from decoherence and unwanted entanglement with other circuit elements, favourable for fault-tolerance.

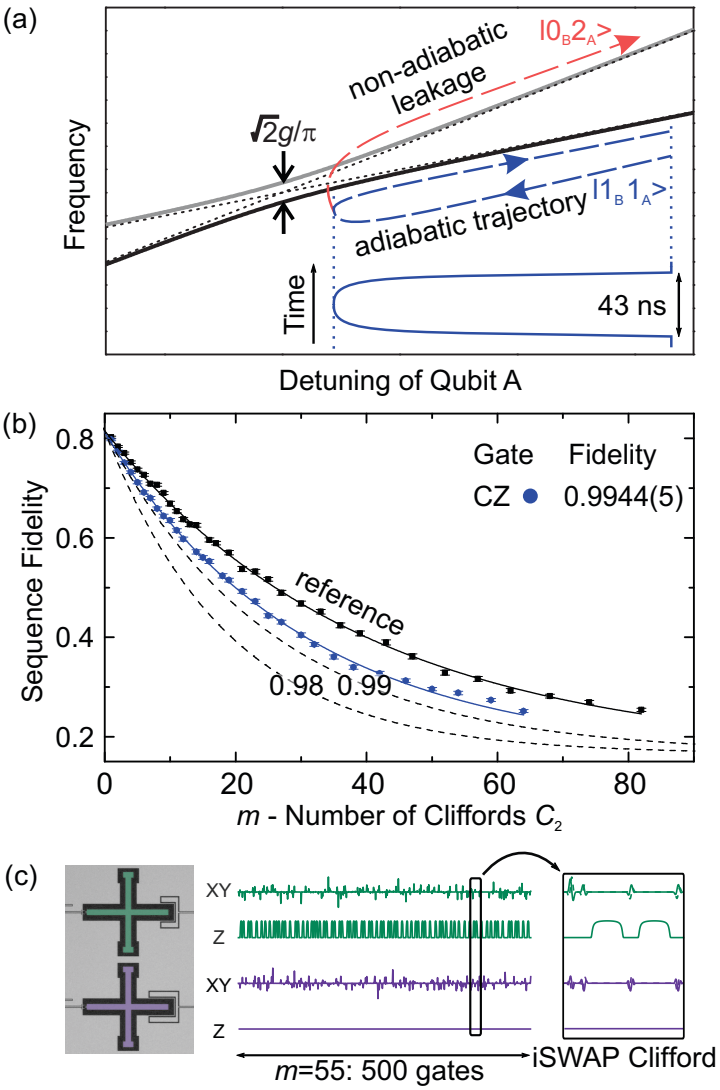


Figure 4.3: **CZ gate physics and randomised benchmarking results.** (a) We use the the $|1_B 1_A\rangle$ and $|0_B 2_A\rangle$ avoided level crossing to implement a high-fidelity CZ gate, with the fast adiabatic tuning of qubit A giving a selective π phase change of $|1_B 1_A\rangle$. The energy level diagram shows qubit A approaching and leaving the avoided level crossing from above (top, blue dashed line), following a fast (43 ns) yet effectively adiabatic trajectory (bottom, solid blue line). Unwanted state leakage from $|1_B 1_A\rangle$ to $|0_B 2_A\rangle$ (red dashed line) is minimised by adjusting the trajectory. (b) Randomised benchmarking data ($k = 100$) of the CZ gate for the qubit pair Q_2 and Q_3, using the two-qubit Clifford group C_2 (Appendix B); reference data in black ($r_{\text{ref}} = 0.0189$), interleaved in blue ($r_{C_2+\text{CZ}} = 0.0244$). Dashed lines indicate the thresholds for gate fidelities of 0.98 and 0.99. We find a CZ gate fidelity of 0.9944 ± 0.0005 (uncertainty from bootstrapping). (c) Coherent microwave (XY) and frequency (Z) control of the quantum state while performing a complex two-qubit algorithm; the sequence contains over 500 gates, corresponding to the characteristic reference decay of $m = 55$, and is over 7 μs long. The right panel shows an example Clifford from the iSWAP class, comprised of single qubit rotations and two CZ gates (Appendix B).

The benchmarking results of the CZ gate are shown in Fig. 4.3b. Similar to the single-qubit case, we generate sequences of two-qubit Cliffords to produce a reference curve, then interleave the CZ gate to extract the fidelity. An example pulse sequence for an $m = 55$ Clifford sequence is shown in Fig. 4.3c. We find a CZ gate fidelity of up to 99.44 $\pm$ 0.05 %, consistent with the average error per Clifford (Appendix B). We find fidelities between 99.0-99.44% on all four pairs of nearest-neighbour qubits (Appendix B). This comprises a clear demonstration of high-fidelity single- and two-qubit gates in a multi-qubit Josephson quantum processor. The two-qubit gate fidelity compares well with the highest values reported for other mature quantum systems: For nuclear magnetic resonance and ion traps, entangling gate fidelites are as high as 99.5% and 99.3% [**?** **?**]. Importantly, we have verified by simulation that the experimentally obtained gate fidelities are at the threshold for surface code quantum error correction, see Appendix B.

We are optimistic that we can improve upon these gate fidelities with modest effort. The CZ gate fidelity is limited by three error mechanisms: Decoherence (55% of the total error), control error (24%), and state leakage (21%), see Appendix B. Decoherence can be suppressed with enhanced materials and optimised fabrication [**?** **?**]. Imperfections in control arise primarily from reflections and stray inductances in wiring, and can be improved using conventional microwave techniques. Given the adiabatic nature of the CZ gate, 2-state leakage can be suppressed by slightly increasing the gate time [**?**].

4.5 Qubit architecture and gate design

We demonstrate a qubit achitecture for the Xmon that is a marked deviation from traditional transmon coupling schemes. In the cannonical cQED architecture, qubits are coupled through a linear resonator acting as a quantum bus [**? ? ?**]. The purpose of this bus is twofold, to allow coupling between qubits over long distances and reduce the entangling rate when idling. There are two modes of operation for coupling qubits in this manner: virtually or directly. In virtual coupling, qubits are near resonance and the bus frequency is far detuned, thus interactions are realized through virtual occupation of bus states. This has the advantage of reducing population in non-computational bus states, however modest virtual qubit-qubit coupling requires a large qubit-bus coupling. In the direct method, a two-qubit gate requires three steps. 1) The first qubit is put on resonance with the bus to swap in population then moved away 2) the second qubit is brought near resonance with the bus to interact 3) the population is swapped back into the first qubit [**? ?**] (see Fig. 4.4b).

In initial multi-qubit Xmon devices, we used a direct qubit-bus-qubit coupling scheme. We chose this over a virtual scheme as achieving large qubit-bus coupling with Xmons is demanding due to the qubit-bus impedance mismatch. This scheme presented a challenge: a qubit-qubit gate requires three multi-device operations, two qubit-bus swaps and one bus-qubit gate. This lengthens the time required to perform the gate, increasing decoherence error. Additionally, imperfect swaps can cause population to remain in the non-computation bus states after interaction. Our attempts to use this scheme achieved a maximum fidelity of 95%, measured with a non-Clifford primitive of randomized benchmarking, see Fig. 4.4.

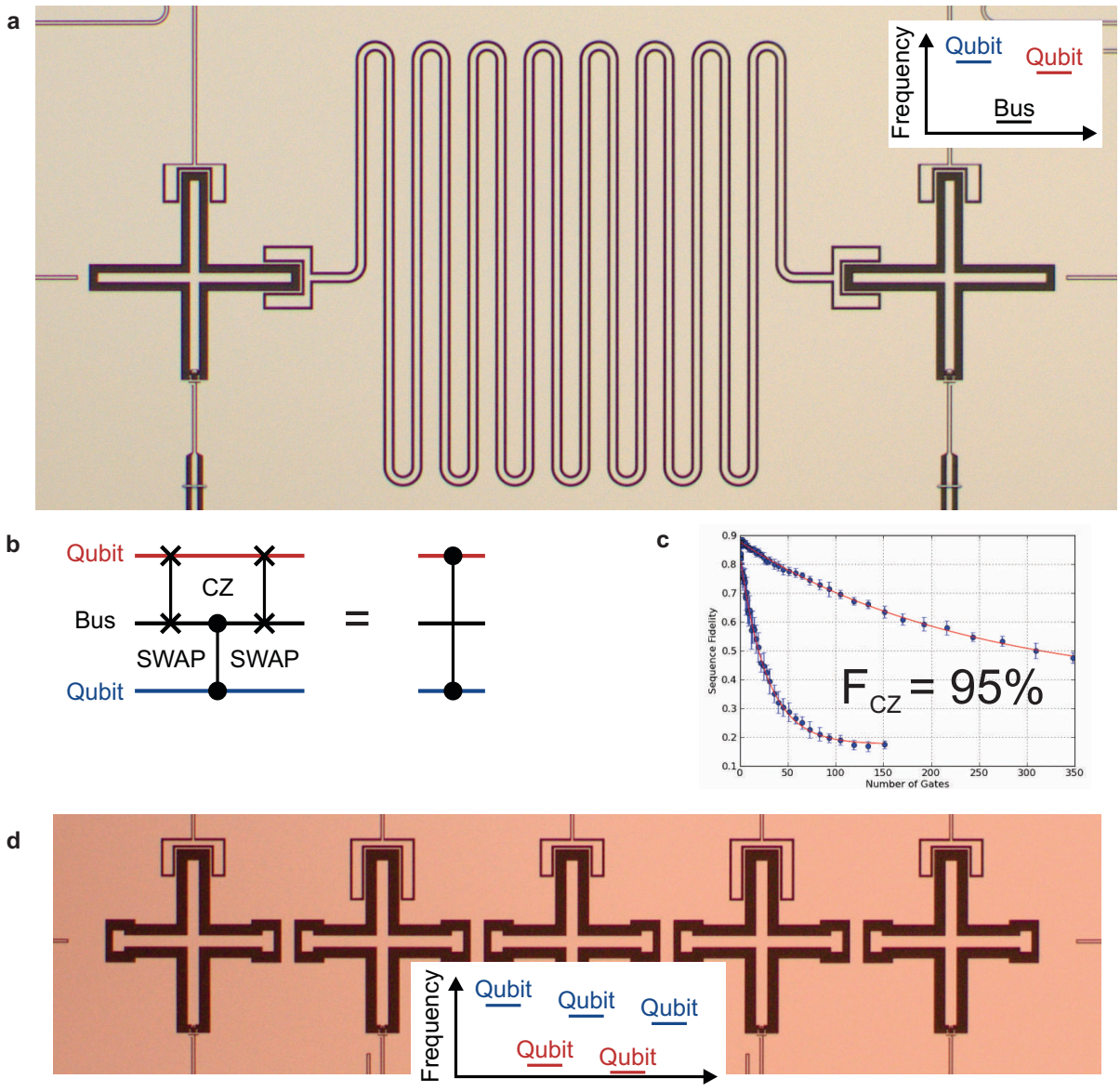


Figure 4.4: **Xmons coupled through quantum bus.** (a) Device micrograph of two Xmon qubits coupled via a resonator acting as a quantum bus. Energy level diagram in inset. (b) Circuit diagram of a qubit-qubit controlled-phase gate, involving two qubit-bus swaps and one qubit-bus controlled-phase gate. (c) Pauli-based randomized benchmarking data for controlled-phase gate, with fidelity of 95 %. (d) Device micrograph of direct coupling scheme, without quantum bus. Energy level diagram in inset.

Rather than improve on the direct coupling scheme, or move to a virtual coupling scheme, we chose to remove the bus resonator entirely. This is possible since a two-dimensional array of qubits, as in the surface code, does not require long distance coupling. Additionally, idling qubit-qubit interactions can be acceptably turned off with frequency detuning the qubits instead of inserting a bus resonator (see Appendix B for qubit-qubit errors arising from off-resonant qubit-qubit interactions, and Chapter 7 for a deeper understanding of these errors). This has the advantage of simplifying the circuit by reducing unnecessary elements, and removing non-computational states which can be an error mechanism. Furthermore, defects in the spectrum can be avoided as multi-qubit gates can be realized at a variety of frequencies, as opposed to requiring direct interaction with the static-frequency bus.

4.6 GHZ states

We showcase the modularity of this set of quantum logic gates by constructing a maximally-entangled GHZ state across all five qubits in our processor, as shown in Fig. 4.5a. The N-qubit GHZ state $|\mathrm{GHZ}\rangle = (|0\rangle^{\otimes N} + |1\rangle^{\otimes N})/\sqrt{2}$ is constructed with single and two-qubit gates, using simultaneous control and readout of all qubits. This algorithm is shown in Fig. 4.5b, the state is assembled by entangling one additional qubit at a time. The algorithm is highly sensitive to control error and decoherence on any computational element. We fully characterise the Bell and GHZ states by using quantum state tomography [**?**], where quadratic maximum likelihood estimation is used to extract each density matrix (ρ) from the measurement data, while satisfying the physical constraints that ρ be Hermitian, unit trace, and positive semi-

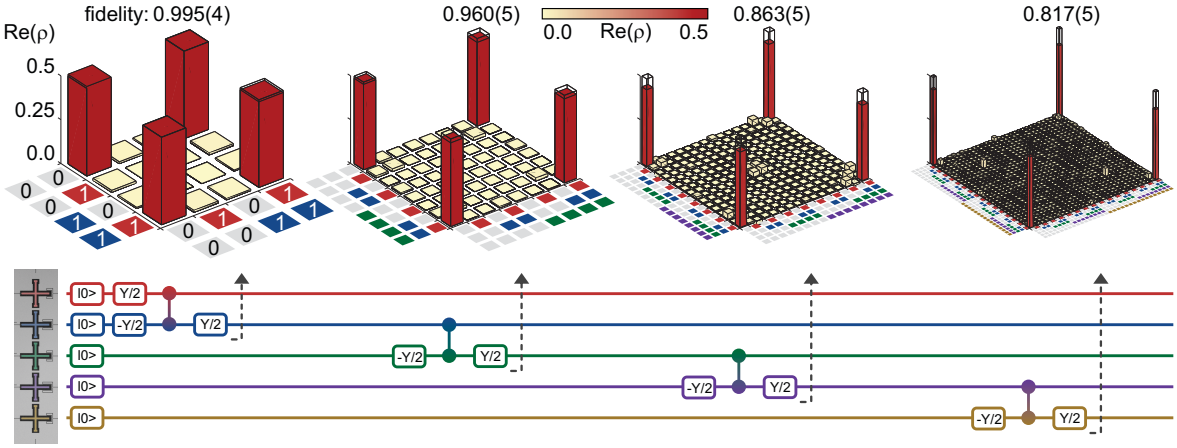


Figure 4.5: **Quantum state tomography and generation of the GHZ state.** Top row: Real part of the density matrix ρ for the $N = 2$ Bell state and the $N = 3$, 4 and 5 GHZ states, measured by quantum state tomography. Ideal density matrix elements are transparent, with value 0.5 at the four corners. Bottom row: Algorithm used to construct the states. See Appendix B for Im(ρ), the Pauli operator representation, and the full gate sequence, which includes Hahn spin-echo pulses.

definite (Appendix B). The density matrices are plotted in the traditional cityscape style, and show significant elements only at the ideal locations. We find state fidelities $\mathrm{Tr}\left(\rho_{\mathrm{ideal}}\rho\right)$ of 99.5 ± 0.4 %, 96.0 ± 0.5 %, 86.3 ± 0.5 % and 81.7 ± 0.5 % for the $N = 2$ Bell state and $N = 3, 4, 5$ GHZ states. A GHZ state fidelity over 50 % satisfies the criterion for genuine entanglement [**?**]. It is interesting to note that the ratio of the off-diagonal to diagonal amplitudes $|\rho_{|0\rangle^{\otimes N},|1\rangle^{\otimes N}}|^2/\rho_{|0\rangle^{\otimes N},|0\rangle^{\otimes N}}\rho_{|1\rangle^{\otimes N},|1\rangle^{\otimes N}}$ have the values 0.99, 0.98, 0.99 and 0.99, suggesting that dephasing is small and/or uncorrelated. The five-qubit GHZ state is the largest multi-qubit entanglement demonstrated to date in the solid state [**? ?**], with state fidelity similar to results obtained in ion traps [**?**]. This demonstrates that complex quantum states can be constructed with high fidelity in a modular fashion, highlighting the potential for more intricate algorithms on this multi-purpose quantum processor.

4.7 Conclusion

We have shown single and two-qubit gates with fidelities at the fault-tolerant threshold for the surface code in an integrated circuit quantum processor. With this demonstration, Josephson quantum devices are now poised to explore fault-tolerant, multi-qubit computing. Extending the linear array of qubits to larger qubit numbers is straightforward, and generating a two-dimensional grid of qubits appears to be mostly a (significant) engineering challenge in adding more qubits, wiring and readout, while maintaining coherence and gate fidelity (Appendix B). In a separate experiment, we have demonstrated qubit state measurement with 99% measurement fidelity in 140 ns [**?**], in a design that can be seamlessly integrated with this architecture. The combination of high-fidelity logic, a multi-qubit architecture, and fast and accurate qubit readout provides the essential ingredients for a Josephson surface code quantum computer.

Chapter 5

Optimal quantum control using randomized benchmarking[1]

5.1 Introduction

Quantum information is stored in continuous amplitudes and phases, so quantum control must be precise to achieve the desired state [**?**]. Achieving control with high fidelity lies at the heart of enabling fault-tolerant quantum computing [**? ?**]. With gate fidelities approaching the fault-tolerant threshold [**? ? ?**], characterizing and reducing the remnant error becomes increasingly challenging. Quantum process tomography can completely characterize a gate, decomposing a process into Pauli or Kraus operators [**? ?**]. However, improving gates is complicated: gate parameters map non-intuitively onto the process matrix, and state preparation and measurement

[1]The following was published as: "Optimal quantum control using randomized benchmarking", Julian Kelly, Rami Barends, et al. *Phys. Rev. Lett.* 112, 240504 (2014).

errors (SPAM) can be confused with process errors.

Here, we present a different approach to achieve high fidelity gates. We use Clifford-based randomized benchmarking (RB) [**?** **?**] to map gate errors onto control parameters, and feed this back to optimize gates. The method is fast and scales to arbitrary precision as the sensitivity to fractional error is independent of gate fidelity (Appendix C); tuning up even higher fidelity gates should be possible. We apply it to general quantum control problems, such as gate optimization, gate bleed-through [**?**] and crosstalk. In particular, we demonstrate closed-loop optimization with nonorthogonal parameters in a real, noisy quantum system. As RB is platform-independent, our approach is in principle applicable to a variety of quantum systems.

In standard RB, gates are characterized by measuring the fidelities of sequences with varying lengths. We experimentally show that optimizing the *sequence* fidelity at fixed length improves the *gate* fidelity. We call this approach – using the sequence fidelity from randomized benchmarking as a fitness metric for gate performance – optimized randomized benchmarking for immediate tune-up (ORBIT).

As a testbed, we use a five qubit (Q_0-Q_4) superconducting system (Appendix C) based on the Xmon transmon design [**?**]. Here, XY control is achieved with microwave pulses and Z control with DC current pulses that modulate the qubit frequency. Qubits are coupled capacitively. Qubit frequencies are tuned between $f_{10} = 4$ and 6 GHz, and qubit nonlinearities $\Delta/2\pi$ are around -220 MHz. This device is an ideal platform for optimizing for small errors, as we have obtained high fidelity single- and two-qubit gates on this device (see Ref. [**?**] and

Appendix C for details on device and setup). We use the same device here.

5.2 Superconducting qubits and pulse shaping

High-fidelity gates require errors from both decoherence and control to be low. There is a fundamental tension between coherence and control in a quantum gate: gates must be fast to minimize decoherence error, but fast gates can introduce control error due to imperfect waveforms or fundamental control constraints. In platforms such as ion traps which have relatively long coherence times, high-fidelity gates can be constructed from slow, simple waveforms with few parameters. However, even with the coherence times of the Xmon qubit of chapter 3, superconducting qubits do not have this luxury. Instead, we introduce sophisticated pulses with shaping parameters so that gates can be performed as quickly as possible, with minimal control error.

This begs the question: how does one optimize a gate with many, and in general non-orthogonal, parameters? The simplest method is to specially design an experiment for each gate parameter. However, this method can fail if parameters are non-orthogonal, or if there is no simple metric to measure error in a parameter. Ideally, we would introduce a single fitness function that can evaluate a set of parameters simultaneously. ORBIT was developed to do just that: optimize gates based on the fidelity measured from randomized benchmarking. This work was critical in achieving the fidelities demonstrated in Chapter 4.

5.3 Optimization of single parameters

We start with a simple test case where we optimize a single-qubit 90 degree rotation about the X-axis in the Bloch sphere representation (X/2 gate). This gate is implemented by a microwave pulse with a cosine envelope (Fig. 5.1a inset) centered around frequency f with amplitude A. As the Xmon transmon qubit is a multilevel system, we apply a quadrature correction term with DRAG coefficient α to minimize leakage to higher levels [**?** **?** **?**]. First, we determine the gate fidelity using RB, then measure how control errors affect the fidelity of sequences.

In Clifford-based RB, random Clifford rotations are inserted between the gate under test to ensure that it is applied to a representative set of states. The single-qubit Clifford gates are the group of rotations that map between the two polar and four equally spaced equator states on the Bloch sphere, and are able to generate a sufficient set of states to remove bias from gate error. To quantify the X/2 fidelity, we first measure a reference curve by applying many sequences of random Cliffords, appended by recovery Cliffords C_r that make the ideal operation the identity. As we initialize the qubit in the ground state, the ground state population becomes the sequence fidelity. Randomization makes the sequence fidelity follow an exponential decay from the accumulation of gate-aspecific errors as $Ap_{\mathrm{ref}}^m + B$, with gate errors captured in the characteristic scale p_{ref} (see Fig. 5.1a). The SPAM errors affect A and B, but not the rate of the decay. Individual gate fidelities are evaluated by interleaving a specific gate between Cliffords, generating a decay curve with scale p_{gate}. By subtracting away the reference curve, we get the gate error $r_{\mathrm{gate}} = (1 - p_{\mathrm{gate}}/p_{\mathrm{ref}})(d-1)/d$ [**?**], with $d = 2^n$ a function of the number of qubits n; here $n = 1$. Each point in m is an average of the fidelity of k different random sequences.

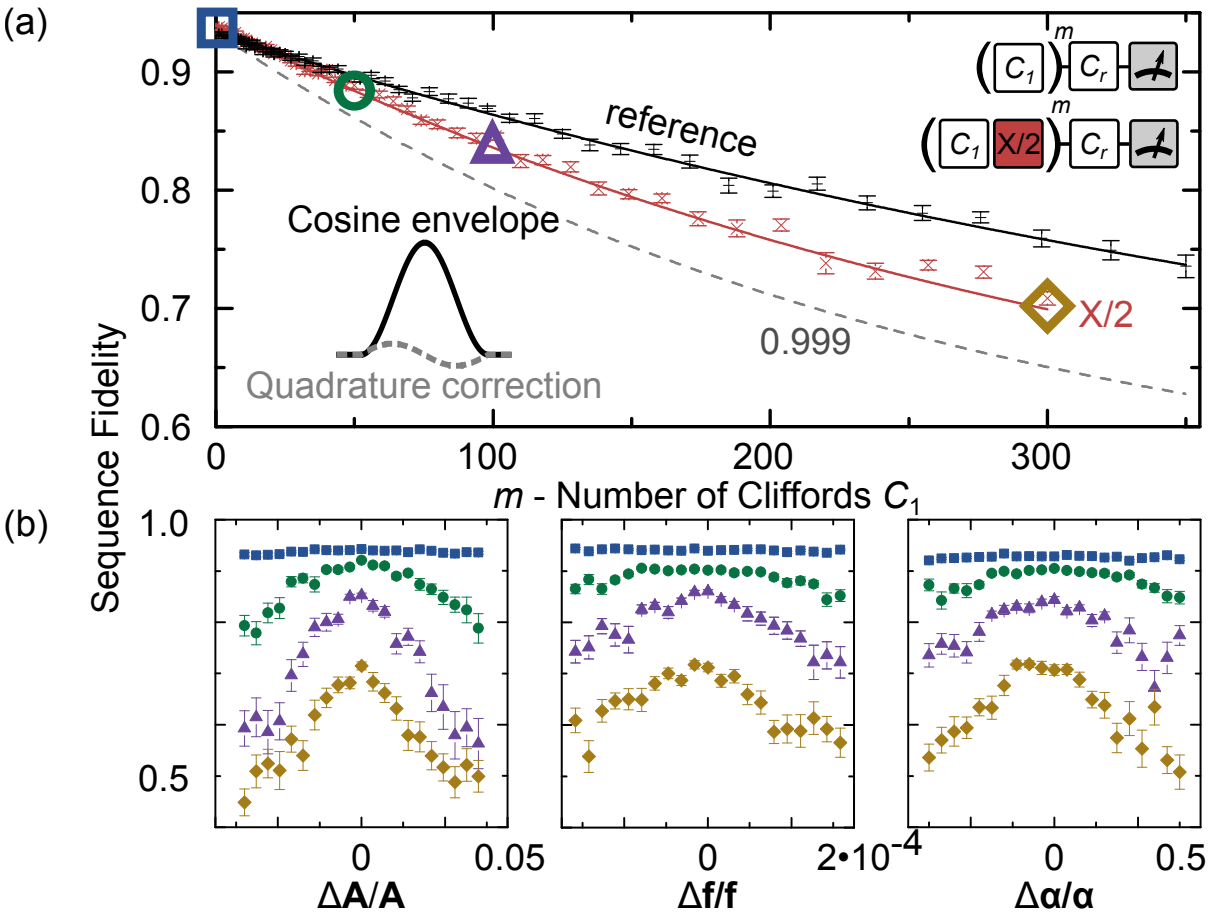


Figure 5.1: (Color online) (a) Single qubit randomized benchmarking ($k = 40$) for Q_2 at $f_{10} = 5.5223$ GHz. The reference experiment uses sequences of random Cliffords only (black plus); a X/2 gate is tested by interleaving it with random Cliffords (red x). The X/2 gate has a cosine envelope pulse shape and quadrature correction (inset). (b) Sequence fidelities versus parameters: the pulse amplitude A, frequency f, and coefficient α ($k = 20$). The sequences are measured at m =1 (square), 50 (circle), 100 (triangle), and 300 (diamond).

We find the fidelity of this X/2 gate to be 0.9995 ($k = 40$).

For the data in Fig. 5.1b, we set $m = 1, 50, 100, 300$ and measure the sequence fidelity as we vary each of the gate parameters from their optimum. As expected, we find that longer length sequences drop more rapidly in fidelity away from the maximum, indicating an increased sensitivity to gate error with sequence length. It is this feature that opens a viable route to optimizing arbitrarily high fidelity gates: sensitivity can be maintained by doubling m when the error is halved (Appendix C).

In the rest of this Chapter, we demonstrate that ORBIT is applicable to a variety of non-trivial parameterized tune-up problems, such as entangling gate optimization with non-orthogonal parameters, improving waveform control for reducing gate bleed-through, and minimizing crosstalk in a multi-qubit system. We emphasize that these applications are issues of prime importance to high fidelity and scaling up to larger qubit systems [**?**].

5.4 Optimizing the eight parameter controlled-phase gate

We start by applying ORBIT to a controlled-phase (CZ) entangling gate with qubits Q_2 and Q_3 that have $g/2\pi = 30$ MHz coupling, as described in Ref. [**?** **?**]. With the addition of many non-orthogonal gate parameters and a larger Hilbert space, this is a significant increase in complexity compared to the X/2 gate. The CZ gate is performed by moving a qubit along an adiabatic trajectory in frequency [**?**] (see inset Fig. 5.2a) which brings the $|11\rangle$ and $|02\rangle$ avoided level crossing near resonance, generating a conditional phase. The fidelity of this gate is sensitive to the frequency trajectory, as deviations from the ideal can cause a conditional

phase other than π as well as non-adiabatic leakage errors to $|02\rangle$. The gate depends on eight parameters that follow straightforwardly from theory (see Ref. [**?**]).

The direct mapping that ORBIT provides between the control parameters and gate fidelity allows for automated optimization. Here, we used the Nelder-Mead algorithm for closed-loop control [**?**]. As a metric, we use sequences ($m = 30$) composed of gates from the two-qubit Clifford group C_2, generated with an average of 8.25 single qubit gates and 1.5 CZ gates per Clifford [**?**]. As single qubit gates have a substantially higher fidelity, the CZ error is a significant contribution to the Clifford error, making the reference fidelity a metric for CZ gate fidelity. Figure 5.2a shows the reference curves before (blue squares) and after (red circles) optimization. The average error per Clifford was reduced from $r = 0.0361$ to $r = 0.0188$. Taking into account the average number of single and CZ gates in a two-qubit Clifford, these values are consistent with a CZ gate fidelity improvement from 0.984 to 0.993 (see Appendix C for interleaved data). Figure 5.2b shows the evolution of the sequence fidelity versus number of evaluations, starting with the blue square; it initially varies strongly with small parameter changes, underlining the sensitivity of this method, and eventually converges on optimal parameters (red circle). The inset of Fig. 5.2a shows the small change in waveform shape (up to 5 MHz in magnitude) that improves fidelity.

Figure 5.2 illustrates the advantages of this approach. First, we can identify and remedy small errors in an environment with noise; we optimize parameters to where gate errors are no longer dominated by control imperfections (see Ref. [**?**] for a representative error budget for a similar experiment). Second, our approach is fast: the total number of measurements is 18000

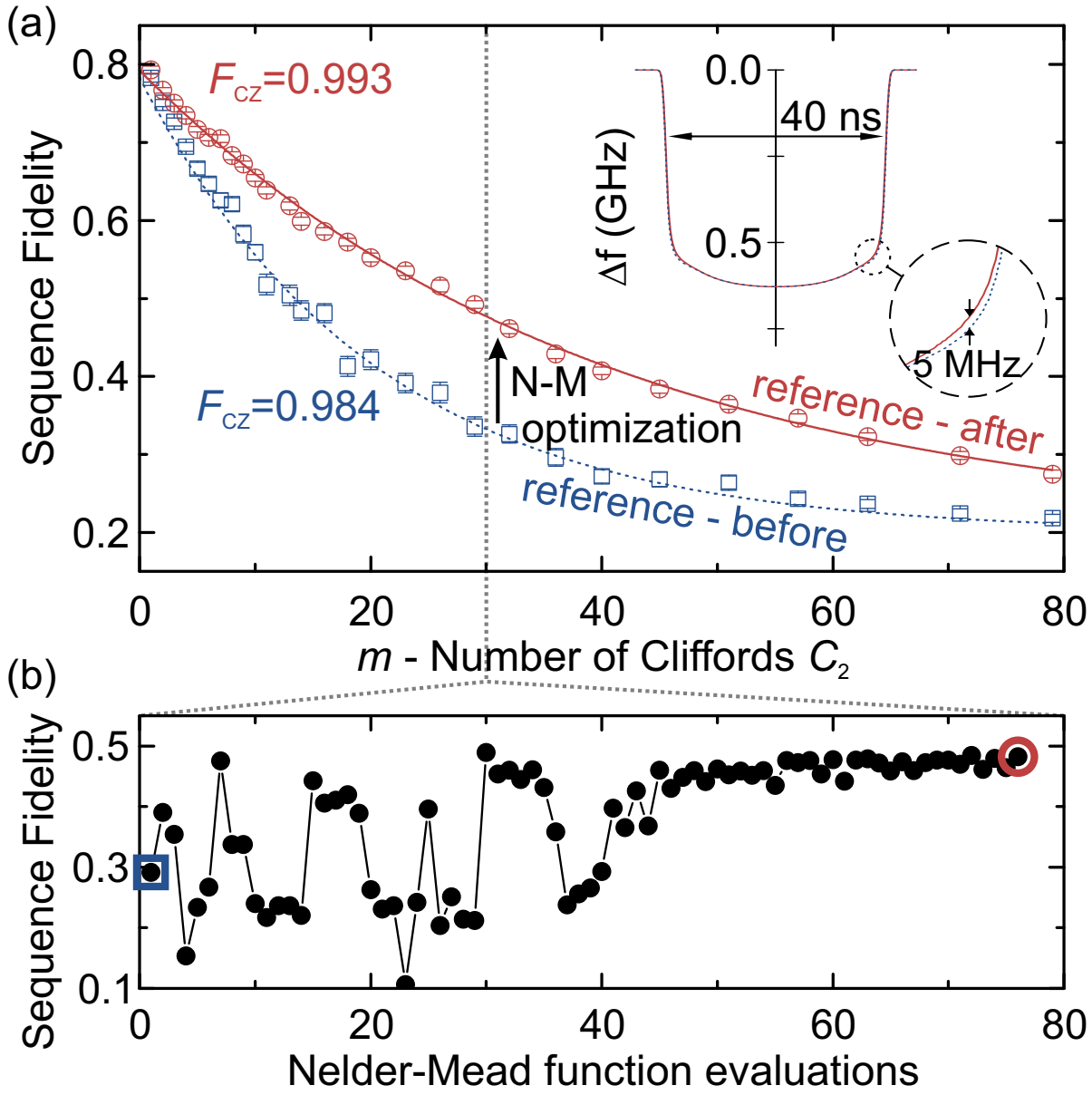


Figure 5.2: (Color online). Optimizing the fidelity of a two-qubit CZ gate. (inset a) One qubit undergoes an effective adiabatic trajectory in frequency that brings the $|11\rangle$ and $|02\rangle$ near resonance, producing a conditional phase. Q_2 idles at $f_{10} = 5.5223$ GHz, and Q_3 idles at $f_{10} = 4.6639$ GHz. (a) The sequence fidelity of the reference curve versus number of two-qubit Cliffords before (blue squares) and after (red circles) optimization ($k = 50$). This optimization has shifted the shoulder of the trajectory by up to 5 MHz (inset). (b) The change in sequence fidelity at $m = 30$ versus Nelder-Mead function evaluations ($k = 20$), starting at a fidelity of 0.3 (blue square), and converging on a sequence fidelity of 0.5 (red circle). The fidelity of the CZ improved from $F_{\mathrm{CZ}} = 0.984$ to $F_{\mathrm{CZ}} = 0.993$, measured using interleaved RB (Appendix C).

(k = 20 sequences, 900 repetitions each), which can be performed in 2 seconds with our system. Third, the optimization is model-free, which is a powerful tool as the system Hamiltonian is not always known to high precision. We believe this will be critical to improving gates beyond current fidelities.

We have used the Nelder-Mead algorithm with ORBIT for automated tune-up as it is a gradient-free method, and therefore less sensitive to noise. While here we used Nelder-Mead for "last-mile" optimization – where gate parameters are initially near the global optimum - any algorithm which uses a fitness metric could be used with ORBIT. Possible applications lie in implementing model-free gates, such as with numerical optimal control [**?** **?**], where pulses are discretized into pixels. This technique can be used on the full Hamiltonian without approximation, can optimize for robustness against noise or experimental parameters, and can generate gates as fast as the "quantum speed limit" [**?**]. Experimentally implementing such gates is hindered by differences between the modeled and actual system Hamiltonians. ORBIT could be a bridge by providing a fitness metric for the closed-loop approaches such as Ad-HOC (see Ref. [**?**]).

5.5 Optimizing gate bleed-through

We now use ORBIT to minimize gate bleed-through; this is a particularly harmful problem because it causes gate-specific errors on potentially many subsequent gates. Gate bleed-through occurs when the mechanism for implementing a gate is not adequately turned off at the end. Physical mechanisms include reflections of control pulses, stray inductance in control lines, and

amplifier slew rates for microwave systems. Gate bleed-through is challenging to characterize and correct, because the entire time domain response must be optimized. Here, we demonstrate that ORBIT is sensitive to a gate bleed-through mechanism, allowing for optimization.

We reduce gate bleed-through from a detune operation which is implemented using a square step pulse on the qubit frequency control line of Q_2. The qubit is detuned for 35 ns by -0.37 GHz, acquiring a single-qubit phase $\phi = 13 \cdot 2\pi$. These current pulses can detune the qubit during subsequent gates if not properly leveled, as illustrated in the top inset of Fig. 5.3a. In the bottom inset, we measure deviations $\delta\phi(t)$ from the ideal acquired qubit phase before and after correction. We compensate the waveform for stray inductances and reflections in the line by applying an inverse transfer function with two poles, expressed in terms of the step response: $\Theta'(t) = \Theta(t)[1 + \sum_i a_i \exp(-\gamma_i t)]$, with $\Theta(t)$ the Heaviside step function, and amplitudes a_i and rates γ_i. In Fig. 5.3a, the error of a Clifford plus step pulse is reduced from $r = 0.011$ to $r = 0.003$ by Nelder-Mead optimization. The sequence fidelity and evolution of the parameters a_i, γ_i, and accumulated qubit phase are shown to converge in Fig. 5.3b and Fig. 5.3c. Gate bleed-through is reduced as evidenced in the improved sequence fidelity. Additionally, the remnant qubit phase $\delta\phi(t)$ is markedly flatter after the detuning pulse as the variance in phase is reduced from $22 \cdot 10^{-4}$ to $8 \cdot 10^{-4}$ rad^2 (see the bottom inset of Fig. 5.3a). This demonstrates that gate bleed-through – here arising from imperfect zeroing of Z control – can be minimized without the need for a full time-domain characterization.

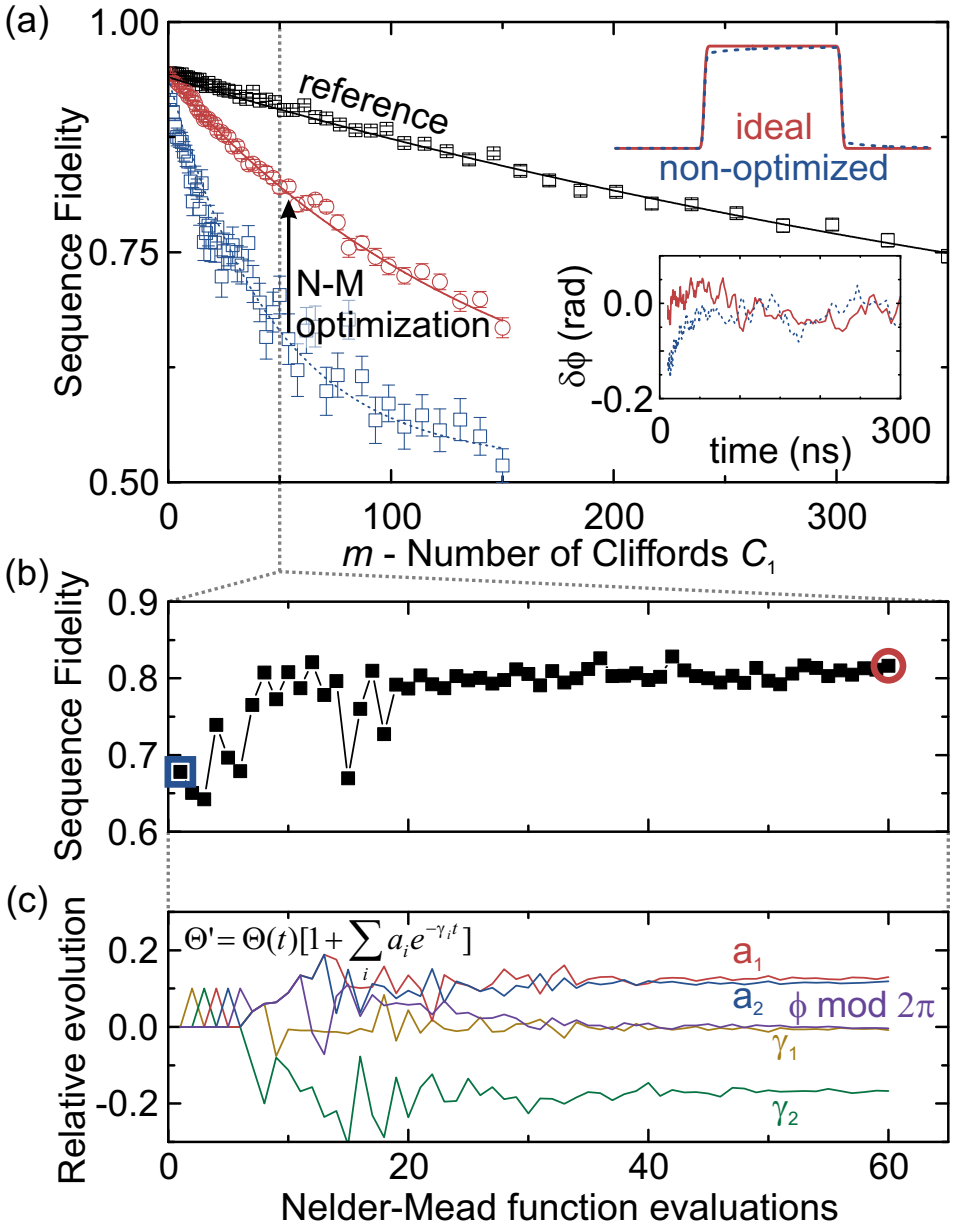


Figure 5.3: (Color online) Reducing gate bleed-through. (top inset a) For rotating the state around the Z axis (Bloch sphere representation), the qubit frequency is detuned by a step pulse, with $t_{\text{gate}} = 35$ ns and large frequency change $\Delta f = -0.37$ GHz which is ideally flat (solid red). Non-idealities in control and wiring bring about a non-trivial deformation of the waveform (dashed blue), causing gate bleed-through. (a) Sequence fidelity *vs* number of Cliffords for the reference (black squares), and interleaved with the step pulse for Q_2 at $f_{10} = 5.5223$ GHz. The error per Clifford plus step pulse is reduced using ORBIT from $r = 0.011$ (blue squares) to $r = 0.003$ (red circles). With this improvement, the remnant qubit phase $\delta\phi(t)$ after the step pulse is notably more flat (bottom inset), determined via quantum state tomography ($k = 40$). (b) Sequence fidelity during the Nelder-Mead algorithm ($k = 30$). (c) Evolution of transfer function parameters, written in terms of the step response Θ'.

5.6 Optimizing control crosstalk

We also apply ORBIT to optimization problems relevant to large systems. One of the greatest challenges in scaling up to larger quantum systems is to maintain addressability over single qubits as control pulses for one qubit can affect others. In our architecture, we minimize control crosstalk by alternating the qubit frequency [**?**]; next-nearest neighbors however are prone to crosstalk due to the smaller frequency difference (see Fig. 5.4a). A difficulty in minimizing crosstalk lies in characterizing its effect on gates. Here, ORBIT provides an elegant solution by mapping errors onto the relevant parameters, through the isolated and simultaneous application of single-qubit Cliffords [**?**].

We start by measuring the reference fidelity curve for qubit labeled Q_2, shown in Fig. 5.4b. From the decay, we find an average error per Clifford of $r_c = 0.001$, consistent with the average single qubit gate fidelity of $F = 0.9995$. The colored regions indicate different ranges in reference fidelity; we use this as a map to infer the gate fidelity from the sequence fidelity. Next, we monitor the sequence fidelity (with $m = 35$) of Q_2 while sending pulses for single-qubit Cliffords down the control line of Q_0. We can ignore the state of Q_0. We vary both the detuning δ and gate length t_{gate} for pulses on the Q_0 line, while keeping the product of gate length and amplitude fixed to mimic control crosstalk. The inferred gate fidelity of qubit Q_2 is shown in Fig. 5.4c. The red regions indicate minimal added error from crosstalk ($< 0.05\%$), while the blue regions show significant increase in error ($> 1\%$). Clear signatures of infidelity appear when crosstalk signals are resonant with the qubit transition frequencies f_{10} or f_{21}, as illustrated in blue in Fig. 5.4a, and fall off with detuning and gate length as expected.

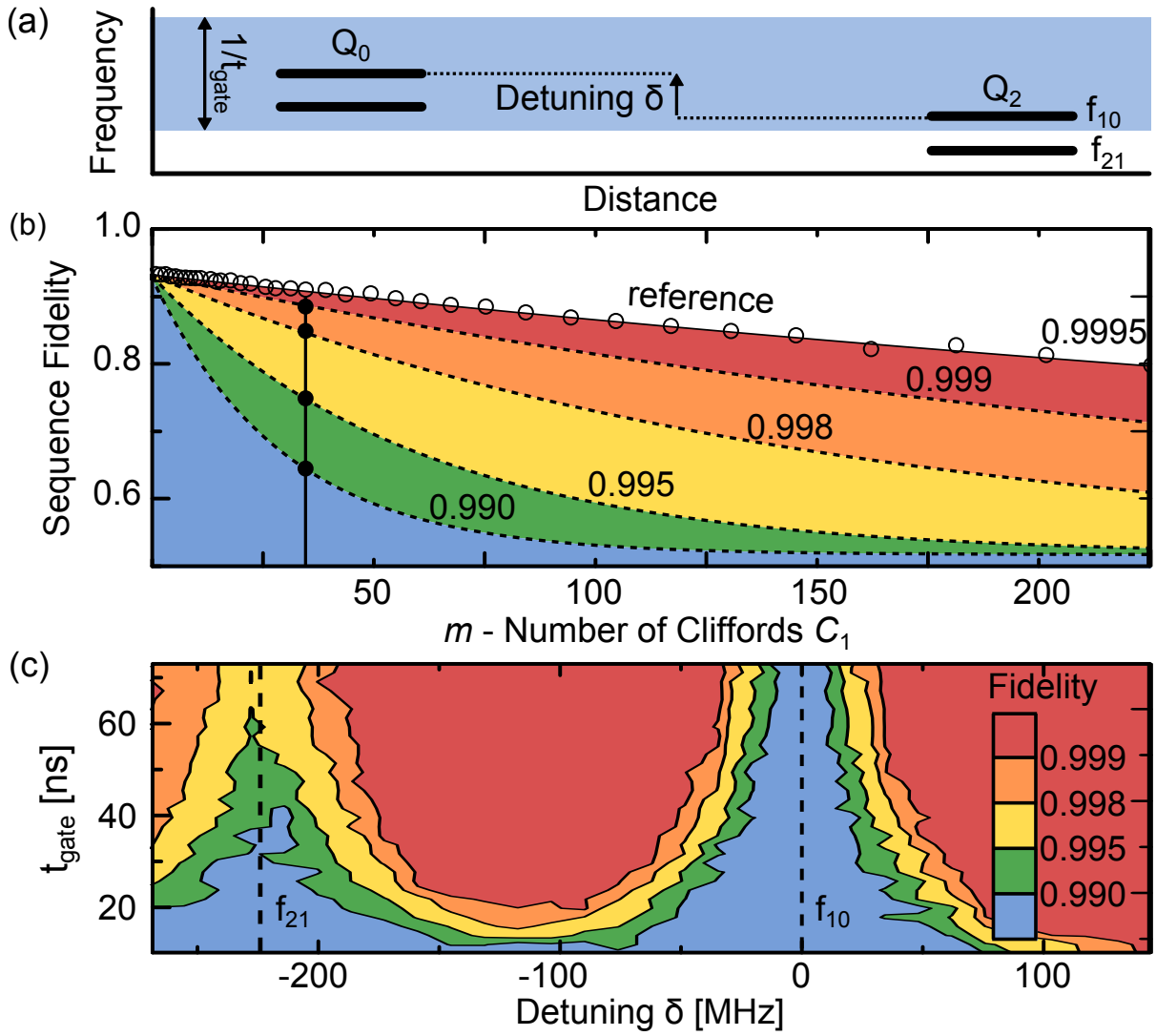


Figure 5.4: (Color online) Mapping control crosstalk. (a) Energy level diagram of the qubits Q_2 and Q_0. Control pulses are applied to the XY lines of Q_2 and Q_0; the latter is swept in detuning δ and gate length $t_{\rm gate}$. Control crosstalk can be exacerbated by small detunings or fast gates. (b) Single qubit benchmarking of Q_2 at $f_{10} = 5.5223$ GHz. The colored regions indicate different ranges of reference fidelity. The vertical cut indicates the m value used to discriminate between regions. (c) The inferred gate fidelity ($m = 35$, $k = 20$) versus detuning and gate length (Appendix C).

The data in Fig. 5.4 demonstrates that ORBIT can provide a map to visualize and optimize control crosstalk in a straightforward manner, without the need to characterize or recalibrate the pulses on qubit Q_0. This technique could in principle also be used for crosstalk reduction methods that reduce spectral power at overlapping frequencies (see Ref. [**? ?**]).

In using ORBIT, we explicitly assume that the cause of sequence decay remains unchanged: the single exponential decay, and SPAM errors captured in parameters A and B, must be consistent. We experimentally find that behavior remains consistent, by comparing standard RB before and after optimization (Fig. 5.2 and Fig. 5.3). This consistency and stable fidelities over many hours suggest that drifts in qubit and control parameters are small. In addition, leakage out of the computational subspace is assumed to penalize sequence fidelity [**?**]. The results show that small leakage errors penalize fidelity for single- and two-qubit gates (Fig. 5.1b,Fig. 5.2). Interestingly, while RB assumes that gate errors are independent of previous gates and Cliffords fully randomize over the computational subspace, we are able to mimimize gate bleed-through and leakage. Clearly, more work needs to be done to fully understand the limitations and capabilities of Clifford-based RB. Because of these subtleties, we emphasize that the reference and interleaved RB data always should be verified for self-consistency [**?**] (Appendix C).

5.7 Conclusion

We have experimentally tested a new approach for optimizing quantum control using randomized benchmarking. This has been shown to be effective for improving single- and two-qubit

gates, minimizing gate bleed-through, and identifying control crosstalk. These experiments are a representative set of control problems for realizing high fidelity gates on large quantum systems. We believe ORBIT can be a generic tool for implementing closed-loop optimization in experimental systems, due to its speed, accuracy and platform independence.

Chapter 6

Rolling quantum dice with a superconducting qubit [1]

One of the key challenges in quantum information is coherently manipulating the quantum state. However, it is an outstanding question whether control can be realized with low error. Only gates from the Clifford group – containing π, $\pi/2$, and Hadamard gates – have been characterized with high accuracy. Here, we show how the Platonic solids enable implementing and characterizing larger gate sets. We find that all gates can be implemented with low error. The results fundamentally imply arbitrary manipulation of the quantum state can be realized with high precision, providing new practical possibilities for designing efficient quantum algorithms.

[1] The following was published as: "Rolling quantum dice with a superconducting qubit", Rami Barends, Julian Kelly, et al. *Phys. Rev. A* 90, 030303(R) (2014).

6.1 Introduction

The Platonic solids have been studied since ancient times for their beauty and symmetry [**?**], and make excellent random number generators [**?**]. Here, we exploit their symmetry for quantum information. Quantum processing would benefit from having a large set of accurate gates to reduce gate count and error [**? ? ?**], yet it is an open question whether arbitrary gates can be implemented with low error – only the restricted group of Clifford gates [**? ?**] has been used with high precision [**? ? ?**]. We use the Platonic solids as a pathway and implement gate sets inspired by the tetrahedron, octahedron, and icosahedron, including gates never previously benchmarked. We achieve low error for all gates. These results illustrate the potential of using unitaries with a fine distribution, and suggest arbitrary rotations can be realized with high accuracy, opening new avenues for performing gates and designing algorithms efficiently.

Recently, major advances have been made in accurately implementing Clifford gates on a variety of platforms. Superconducting qubits, liquid NMR and ion traps have shown single-qubit gate errors ranging from 10^{-3} to 10^{-6} [**? ? ?**], determined via Clifford-based randomized benchmarking (RB). Arbitrary rotations – for quantum chemistry and quantum simulation [**? ? ?**] – can then be approximated by sequences of Cliffords and non-Clifford gates [**?**]. To minimize the accumulation of gate error when implementing these algorithms in present-day quantum processors, one would like to perform such rotations directly. However, high fidelity rotations outside of the Clifford group are yet to be demonstrated. Process verification of non-Clifford gates is a conundrum: Quantum process tomography can be used, but state preparation and measurement error can lead to significant systematic deviations, limiting precision.

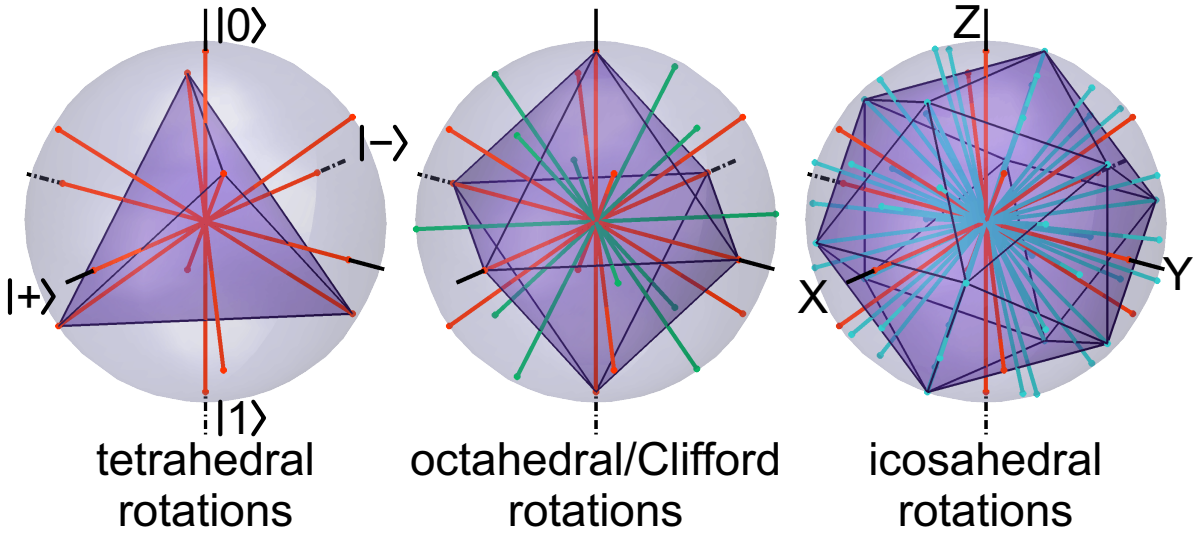


Figure 6.1: (color online) The Platonic solids and their rotational groups. The axes of rotation are of the tetra-, octa- and icosahedral rotational group; the respective Platonic solids are superimposed. The axes are defined by lines intersecting the origin, and a vertex, face center, or midpoint of an edge. The tetrahedral rotational group (orange) is shared among all groups.

Clifford-based RB is insensitive to these errors, but unavailable for gates which fall outside of the Clifford group.

Here, we demonstrate the high fidelity implementation of non-Clifford gates, by using larger rotational groups. Our approach also opens the door to evaluating functions of higher order, and experimentally tests a core premise – that any unitary 2-design is sufficient – of randomized benchmarking, a technique that is becoming a keystone metric in quantum information. A different approach to estimating errors of non-Cliffords was proposed in Ref. [**?**].

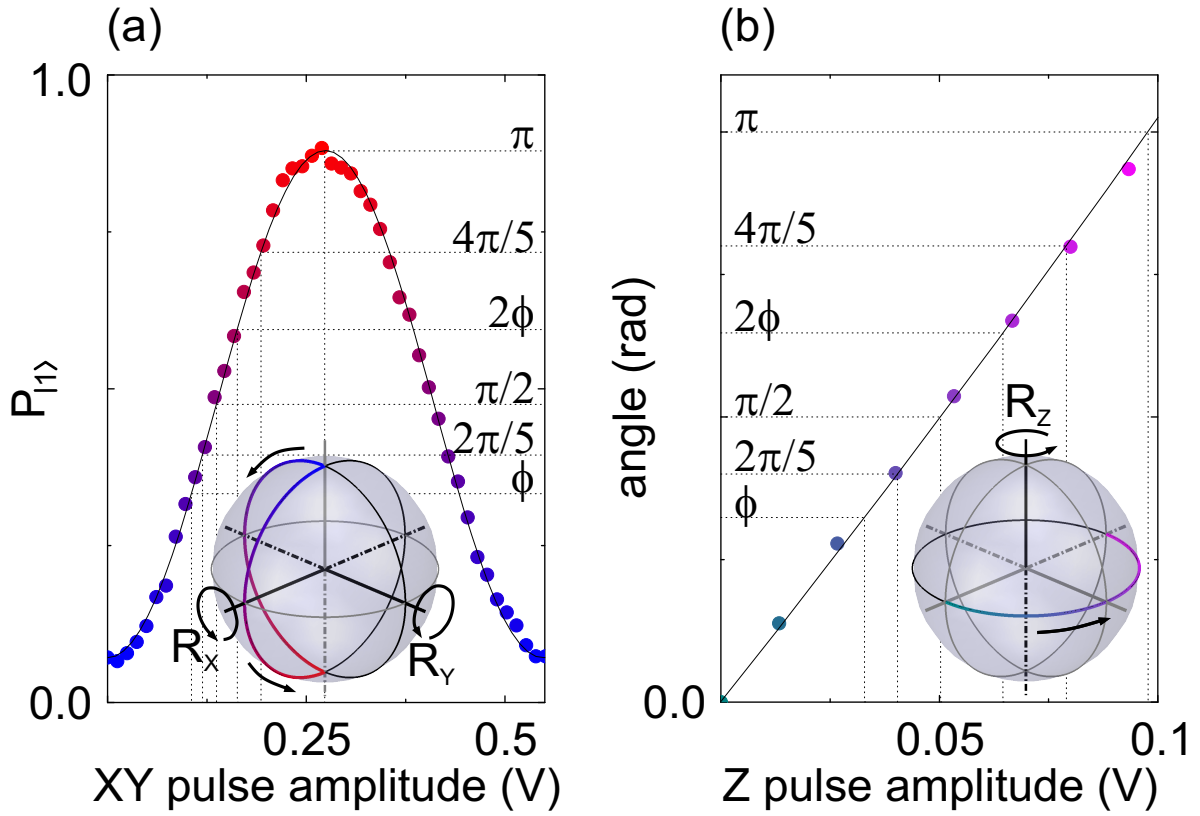


Figure 6.2: (color online) Calibrating the angles of rotation. (a) The excited state probability versus X and Y pulse voltage amplitude on the control board. The amplitudes for the required phases of rotations around the X and Y axes are indicated with dotted lines. The data follow a $\sin^2$ dependence (solid line) on the pulse amplitude, as expected. Data not corrected for measurement fidelity. (b) The phase of the quantum state as a function of Z pulse voltage amplitude, measured using quantum state tomography. Solid line is a fit to the data. For brevity, only the positive angles are shown. Here $\tan\phi = (1+\sqrt{5})/2$. Insets show the trajectories on the Bloch sphere for the X-, Y-, and Z-axis rotations.

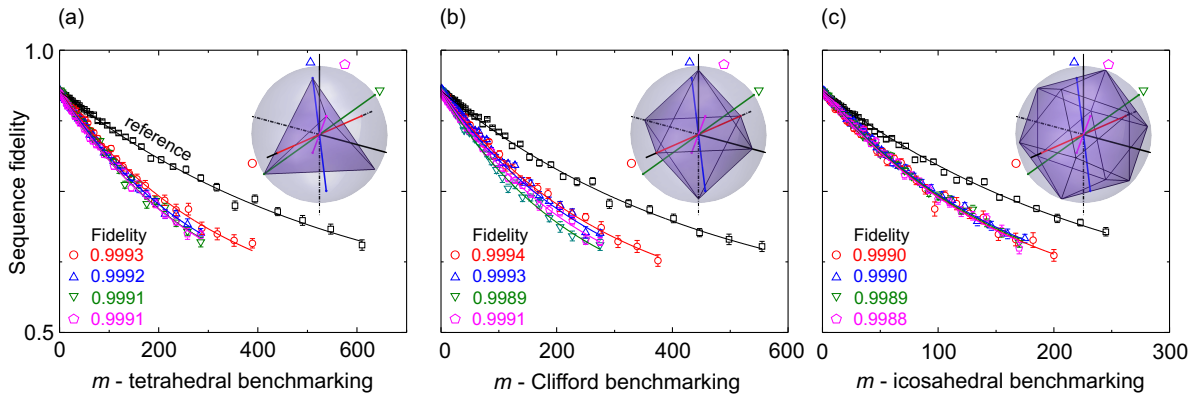


Figure 6.3: (color online) Randomized benchmarking with the (a) tetra-, (b) octa- and (c) icosahedral rotational groups. The sequence fidelities are plotted as a function of m, the number of random rotations or sets of random rotation and interleaved gate. For each m, the fidelity is averaged over $k = 50$ different, random sequences. From fits to the reference curves (black lines) we extract the average error per group rotation of $r_{\mathrm{ref,T}} = 0.0009$, $r_{\mathrm{ref,C}} = 0.0010$, and $r_{\mathrm{ref,I}} = 0.0019$, consistent with an average physical gate fidelity of 0.9995. The rotational groups preserve Platonic solids in Bloch space, the respective solids are shown in the insets. The colored lower curves show the data when interleaving four tetrahedral rotations which are shared among all the three groups, the rotational axes are shown in the insets; the composed gates are X_{π} ($\bigcirc$), $\mathrm{X}_{\pi/2}\,\mathrm{Y}_{\pi/2}$ ($\triangle$), $\mathrm{X}_{-\pi/2}\,\mathrm{Y}_{\pi/2}$ ($\triangledown$), and $\mathrm{Y}_{\pi/2}\,\mathrm{X}_{\pi/2}$ ($\pentagon$). Here, $\mathrm{X}_{\pi/2}\,\mathrm{Y}_{\pi/2}$ denotes the unitary $R_{\mathrm{Y}}(\pi/2) \cdot R_{\mathrm{X}}(\pi/2) = \exp(-i\pi\sigma_{\mathrm{Y}}/4) \cdot \exp(-i\pi\sigma_{\mathrm{X}}/4)$. The gate fidelities are tabulated in the figures, extracted from fits to the data (solid lines). Error bars on the data indicate the standard deviation of the mean. The standard deviations of gate fidelities are typically 10^{-4}

6.2 Unitary 2-designs: tetrahedron, octahedron, and icosahedron

The groups of unitaries we use here are formed by the rotations that preserve the regular tetrahedron, octahedron, and icosahedron – Platonic solids – in the Bloch sphere representation, see Fig. 6.1. These are the rotational subgroups of the tetrahedral, octahedral and icosahedral symmetry groups T_h, O_h and I_h. These rotations exchange faces, amounting to a quantum version of rolling dice (such dice are referred to as d4, d8, and d20), but now in Bloch space. The tetrahedral, octahedral and icosahedral rotational groups have size (order) 12, 24, and 60, respectively. The axes are defined by the lines that intersect the origin, and a face center, vertex, or midpoint of an edge. The angles of rotation around these axes are, respectively, integer multiples of $\{2\pi/3, 2\pi/3, \pi\}$ for the tetrahedral group, $\{2\pi/3, 2\pi/4, \pi\}$ for the octahedral group, and $\{2\pi/3, 2\pi/5, \pi\}$ for the icosahedral group. The tetrahedral rotations (orange axes in Fig. 6.1) are shared among all three groups, enabling comparison experiments. The octahedral rotations form the single-qubit Clifford group. The icosahedral rotations form the most dense group – the icosahedron is the largest of the Platonic solids – allowing for fine unitary control.

For implementing gates from these groups we decompose them into rotations around the X, Y, and Z axes. The tetra- and octahedral groups can be implemented using only $\pi/2$ and π rotations [**?**]. The icosahedral group requires the following rotation angles: $\{\phi, 2\pi/5, \pi/2, 2\phi, 4\pi/5, \pi\}$, with ϕ an irrational angle from $\tan\phi = (1+\sqrt{5})/2$ the golden ratio. The average number of physical gates per tetra-, octa- or icosahedral rotation is $1\frac{3}{4}$, $1\frac{7}{8}$, and $4\frac{4}{15}$, respectively. This

decomposition requires a minimal number of used angles and only one irrational angle.

6.3 Rotational groups

The tetra-, octa-, and icosahedral rotational groups are shown in Tables. 6.1, 6.2 and 6.3.

6.4 Calibration

The rotations are implemented in our superconducting quantum system, the Xmon transmon qubit [**?**]. This qubit combines full, direct axial control with a high level of coherence. Details

Table 6.1: The tetrahedral rotational group, written in terms of the physical microwave gates applied in time. Negative angles are included through opposite rotational axes.

Paulis - π	$2\pi/3$	
I	$X_{\pi/2}$	$Y_{\pi/2}$
X_π	$X_{\pi/2}$	$Y_{-\pi/2}$
Y_π	$X_{-\pi/2}$	$Y_{\pi/2}$
Y_π X_π	$X_{-\pi/2}$	$Y_{-\pi/2}$
	$Y_{\pi/2}$	$X_{\pi/2}$
	$Y_{\pi/2}$	$X_{-\pi/2}$
	$Y_{-\pi/2}$	$X_{\pi/2}$
	$Y_{-\pi/2}$	$X_{-\pi/2}$

Table 6.2: The octahedral rotational group – single qubit Cliffords. The Paulis and $2\pi/3$ rotations form the tetrahedral rotational group.

Paulis - π	$2\pi/3$		$\pi/2$			Hadamard-like - π		
I	$X_{\pi/2}$	$Y_{\pi/2}$	$X_{\pi/2}$			X_π	$Y_{\pi/2}$	
X_π	$X_{\pi/2}$	$Y_{-\pi/2}$	$X_{-\pi/2}$			X_π	$Y_{-\pi/2}$	
Y_π	$X_{-\pi/2}$	$Y_{\pi/2}$	$Y_{\pi/2}$			Y_π	$X_{\pi/2}$	
Y_π X_π	$X_{-\pi/2}$	$Y_{-\pi/2}$	$Y_{-\pi/2}$			Y_π	$X_{-\pi/2}$	
	$Y_{\pi/2}$	$X_{\pi/2}$	$X_{-\pi/2}$	$Y_{\pi/2}$	$X_{\pi/2}$	$X_{\pi/2}$	$Y_{\pi/2}$	$X_{\pi/2}$
	$Y_{\pi/2}$	$X_{-\pi/2}$	$X_{-\pi/2}$	$Y_{-\pi/2}$	$X_{\pi/2}$	$X_{-\pi/2}$	$Y_{\pi/2}$	$X_{-\pi/2}$
	$Y_{-\pi/2}$	$X_{\pi/2}$						
	$Y_{-\pi/2}$	$X_{-\pi/2}$						

of the device used in this experiment can be found in Ref. [**?**]. Rotations around the X and Y axes are achieved by applying microwave pulses. Rotations around the Z axis can be directly performed by detuning the qubit frequency, or by combining X and Y rotations. All control pulses have cosine envelopes, generated by fast (1 Gsample/sec) digital-to-analog converter boards. For XY control we generate both the in-phase and quadrature component and upconvert it to the qubit frequency using quadrature mixing, see Refs. [**?** **?**] for more detail. For calibrating the pulse amplitudes we use the measured probability for X and Y rotations, and for Z rotations the phase as determined using quantum state tomography (Fig. 6.2). We minimize leakage to energy levels above the computational subspace by applying a quadrature correction [**?** **?**]. Subsequently, fine-tuning of the parameters is done through optimized randomized benchmarking for immediate tune-up (ORBIT) [**?**], reducing gate errors by approximately

Table 6.3: The 60 icosahedral rotations, excluding the idle (I). The rotations are ordered based on their angles, and their points of intersection with the icosahedron. The edges and faces contain the Paulis and $2\pi/3$ rotations which overlap with the tetrahedral rotational group. For the edge rotations we have used $R_{\mathrm{X}}(\phi)R_{\mathrm{Y}}(\pi)R_{\mathrm{X}}(-\phi) = R_{\mathrm{X}}(2\phi)R_{\mathrm{Y}}(\pi)$ and $R_{\mathrm{X}}(\phi)R_{\mathrm{Z}}(\pi)R_{\mathrm{X}}(-\phi) = R_{\mathrm{X}}(2\phi)R_{\mathrm{Z}}(\pi)$ to reduce the gate count.

Vertices - $2\pi/5$		
Y_{ϕ}	$\mathrm{X}_{2\pi/5}$	$\mathrm{Y}_{-\phi}$
Y_{ϕ}	$\mathrm{X}_{-2\pi/5}$	$\mathrm{Y}_{-\phi}$
$\mathrm{Y}_{-\phi}$	$\mathrm{X}_{2\pi/5}$	Y_{ϕ}
$\mathrm{Y}_{-\phi}$	$\mathrm{X}_{-2\pi/5}$	Y_{ϕ}
Z_{ϕ}	$\mathrm{Y}_{2\pi/5}$	$\mathrm{Z}_{-\phi}$
Z_{ϕ}	$\mathrm{Y}_{-2\pi/5}$	$\mathrm{Z}_{-\phi}$
$\mathrm{Z}_{-\phi}$	$\mathrm{Y}_{2\pi/5}$	Z_{ϕ}
$\mathrm{Z}_{-\phi}$	$\mathrm{Y}_{-2\pi/5}$	Z_{ϕ}
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	$\mathrm{X}_{-\phi}$
$\mathrm{X}_{-\phi}$	$\mathrm{Z}_{2\pi/5}$	X_{ϕ}
$\mathrm{X}_{-\phi}$	$\mathrm{Z}_{-2\pi/5}$	X_{ϕ}
Vertices - $4\pi/5$		
Y_{ϕ}	$\mathrm{X}_{4\pi/5}$	$\mathrm{Y}_{-\phi}$
Y_{ϕ}	$\mathrm{X}_{-4\pi/5}$	$\mathrm{Y}_{-\phi}$
$\mathrm{Y}_{-\phi}$	$\mathrm{X}_{4\pi/5}$	Y_{ϕ}
$\mathrm{Y}_{-\phi}$	$\mathrm{X}_{-4\pi/5}$	Y_{ϕ}
Z_{ϕ}	$\mathrm{Y}_{4\pi/5}$	$\mathrm{Z}_{-\phi}$
Z_{ϕ}	$\mathrm{Y}_{-4\pi/5}$	$\mathrm{Z}_{-\phi}$
$\mathrm{Z}_{-\phi}$	$\mathrm{Y}_{4\pi/5}$	Z_{ϕ}
$\mathrm{Z}_{-\phi}$	$\mathrm{Y}_{-4\pi/5}$	Z_{ϕ}
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	$\mathrm{X}_{-\phi}$
$\mathrm{X}_{-\phi}$	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}
$\mathrm{X}_{-\phi}$	$\mathrm{Z}_{-4\pi/5}$	X_{ϕ}

Faces - $2\pi/3$							
$\mathrm{X}_{-\pi/2}$	$\mathrm{Y}_{-\pi/2}$						
$\mathrm{Y}_{\pi/2}$	$\mathrm{X}_{\pi/2}$						
X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	X_{ϕ}	$\mathrm{X}_{-\pi/2}$	$\mathrm{Y}_{-\pi/2}$	X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	X_{ϕ}	$\mathrm{Y}_{\pi/2}$	$\mathrm{X}_{\pi/2}$	X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}	$\mathrm{X}_{-\pi/2}$	$\mathrm{Y}_{-\pi/2}$	X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}	$\mathrm{Y}_{\pi/2}$	$\mathrm{X}_{\pi/2}$	X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}
$\mathrm{X}_{-\pi/2}$	$\mathrm{Y}_{\pi/2}$						
$\mathrm{Y}_{-\pi/2}$	$\mathrm{X}_{\pi/2}$						
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	X_{ϕ}	$\mathrm{X}_{-\pi/2}$	$\mathrm{Y}_{-\pi/2}$	X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	X_{ϕ}	$\mathrm{Y}_{\pi/2}$	$\mathrm{X}_{\pi/2}$	X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	X_{ϕ}
$\mathrm{X}_{\pi/2}$	$\mathrm{Y}_{\pi/2}$						
$\mathrm{Y}_{-\pi/2}$	$\mathrm{X}_{-\pi/2}$						
X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	X_{ϕ}	$\mathrm{X}_{\pi/2}$	$\mathrm{Y}_{\pi/2}$	X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	X_{ϕ}	$\mathrm{Y}_{-\pi/2}$	$\mathrm{X}_{-\pi/2}$	X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}	$\mathrm{X}_{\pi/2}$	$\mathrm{Y}_{\pi/2}$	X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{ϕ}	$\mathrm{Y}_{-\pi/2}$	$\mathrm{X}_{-\pi/2}$	X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	X_{ϕ}	$\mathrm{X}_{\pi/2}$	$\mathrm{Y}_{\pi/2}$	X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	X_{ϕ}
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	X_{ϕ}	$\mathrm{Y}_{-\pi/2}$	$\mathrm{X}_{-\pi/2}$	X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	X_{ϕ}
$\mathrm{X}_{\pi/2}$	$\mathrm{Y}_{-\pi/2}$						
$\mathrm{Y}_{\pi/2}$	$\mathrm{X}_{-\pi/2}$						

Edges - π				
X_{π}				
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	X_{π}	$\mathrm{Z}_{-2\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	X_{π}	$\mathrm{Z}_{2\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	X_{π}	$\mathrm{Z}_{-4\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	X_{π}	$\mathrm{Z}_{4\pi/5}$	$\mathrm{X}_{-\phi}$
Y_{π}				
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	Y_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{-2\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	Y_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{2\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	Y_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{-4\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	Y_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{4\pi/5}$	$\mathrm{X}_{-\phi}$
Z_{π}				
X_{ϕ}	$\mathrm{Z}_{2\pi/5}$	Z_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{-2\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-2\pi/5}$	Z_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{2\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{4\pi/5}$	Z_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{-4\pi/5}$	$\mathrm{X}_{-\phi}$
X_{ϕ}	$\mathrm{Z}_{-4\pi/5}$	Z_{π}	$\mathrm{X}_{2\phi}$	$\mathrm{Z}_{4\pi/5}$	$\mathrm{X}_{-\phi}$

10^{-4} [**?**]. The generators of the tetrahedral and octahedral group are fully parameterized by a total of three parameters, and the generators of the icosahedral group by a total of 14 variables (Table 6.4).

6.5 Testing 2-designs

We test the gates using randomized benchmarking [**?** **?** **?** **?** **?**]. In essence, randomized benchmarking is equivalent to randomly rolling the die in Bloch space m times followed by a final rotation that returns it to the starting position, and then measuring the probability of success. One would like to determine the gate error averaged over all possible input states. As the gate error depends quadratically on, for example, any amount of over- or underrotation,

Table 6.4: Generators of the tetra-, octa-, and icosahedral rotational groups, excluding the idle. The generators are listed in terms of shared pulse amplitude parameter. Including the DRAG parameter, we use a total of three parameters to generate the tetra-, and octahedral rotational group (DRAG parameter, pulse amplitude parameter for X_{π} and Y_{π}, pulse amplitude parameter for $\mathrm{X}_{\pi/2}$, $\mathrm{X}_{-\pi/2}$, $\mathrm{Y}_{\pi/2}$, and $\mathrm{Y}_{-\pi/2}$), and a total of 14 parameters to generate the icosahedral rotational group. The duration of the idle and each of the X and Y gates is 12 ns, and each Z gate is 10 ns.

Rotational group	Generators
Tetrahedral, octahedral, and icosahedral	X_{π}, Y_{π} $\mathrm{X}_{\pi/2}$, $\mathrm{X}_{-\pi/2}$, $\mathrm{Y}_{\pi/2}$, $\mathrm{Y}_{-\pi/2}$
Icosahedral	$\mathrm{X}_{2\pi/5}$, $\mathrm{X}_{-2\pi/5}$, $\mathrm{Y}_{2\pi/5}$, $\mathrm{Y}_{-2\pi/5}$ $\mathrm{X}_{4\pi/5}$, $\mathrm{X}_{-4\pi/5}$, $\mathrm{Y}_{4\pi/5}$, $\mathrm{Y}_{-4\pi/5}$ X_{ϕ}, Y_{ϕ}, $\mathrm{X}_{-\phi}$, $\mathrm{Y}_{-\phi}$ $\mathrm{X}_{2\phi}$ $\mathrm{Z}_{2\pi/5}$ $\mathrm{Z}_{-2\pi/5}$ Z_{ϕ} $\mathrm{Z}_{-\phi}$ $\mathrm{Z}_{4\pi/5}$ $\mathrm{Z}_{-4\pi/5}$ Z_{π}

we do not need to evaluate a continuum of input states. The average of a polynomial function of order t over the surface of a sphere can be evaluated exactly using only a finite number of points, such a group of points is a spherical t-design. For the single-qubit case, unitary designs are the group of rotations that can generate spherical designs, mapping between the points [**?**]. Therefore, the rotational group used in randomized benchmarking needs to be a unitary 2-design [**?** **?** **?** **?**]. Moreover, unitary 2-designs depolarize any error in the computational basis. For the single-qubit case, the rotational groups which preserve Platonic solids are the 2-designs [**?**]. There are only three unitary 2-designs as the cube shares the same group as the octahedron (the cube and the octahedron are duals), and the dodecahedron shares the same rotations as the icosahedron (the dodecahedron and the icosahedron are duals). We have thus tested all unitary 2-designs in Bloch space.

Randomized benchmarking with 2-designs is therefore a crucial test of coherent control. The decrease of the probability of success – the sequence fidelity – with increasing sequence length is used to quantify the gate fidelity. We start by measuring a reference curve, using sequences of m random rotations. The sequence fidelity follows $Ap^m + B$, with variables A and B absorbing measurement and initialization errors, and p_{ref} giving the average error per rotation: $r_{\mathrm{ref}} = (1 - p_{\mathrm{ref}})/2$ [**?**]. We then interleave a specific gate with m random rotations, the difference with the reference is a direct measure of the gate error: $r_{\mathrm{gate}} = (1-p_{\mathrm{gate}}/p_{\mathrm{ref}})/2$, the gate fidelity is $F_{\mathrm{gate}} = 1 - r_{\mathrm{gate}}$ [**?**]. At each m, the data is averaged over $k = 50$ random sequences [**?**].

We have performed randomized benchmarking using the tetrahedral, octahedral, and icosa-

hedral rotational groups; the results are shown in Fig. 6.3. As we start by initializing the qubit in the ground state, the sequence fidelity is given by the ground state population after applying the random sequences. The traces follow an exponential decay with increasing m, as expected. We have also interleaved four gates from the tetrahedral group (see insets for the rotational axes). These rotations are shared by all three rotational groups, allowing for a direct comparison between tetra-, octa-, and icosahedral-based randomized benchmarking. We emphasize that the interleaved gates are physically implemented in exactly the same manner.

From the reference traces, we extract an average error per group of rotations of $r_{\mathrm{ref,T}} = 9 \cdot 10^{-4}$, $r_{\mathrm{ref,C}} = 10 \cdot 10^{-4}$, $r_{\mathrm{ref,I}} = 19 \cdot 10^{-4}$. When dividing by $1\frac{3}{4}$, $1\frac{7}{8}$ or $4\frac{4}{15}$, these numbers consistently point to an average error of $5 \cdot 10^{-4}$ per physical gate (single decomposed rotation around the X, Y, or Z axis). The gate errors are dominated by decoherence [**?**]. The extracted fidelities for the interleaved gates are tabulated in Fig. 6.3. The reference error per gate, as well as the errors for the interleaved gates, are consistent with previous measurements [**?**], where the average physical gate fidelity lies at 0.9994. In addition, the mean difference in error of the interleaved gates is below $2 \cdot 10^{-4}$ [**?**], verifying that any of the groups can be used for randomized benchmarking.

6.6 Benchmarking non-Clifford rotations

With icosahedral randomized benchmarking shown to be a viable method for determining gate fidelity, we can now benchmark gates outside of the Clifford group, as shown in Fig. 6.4. We chose three composite gates, which are implemented using three, six, or eight physical gates.

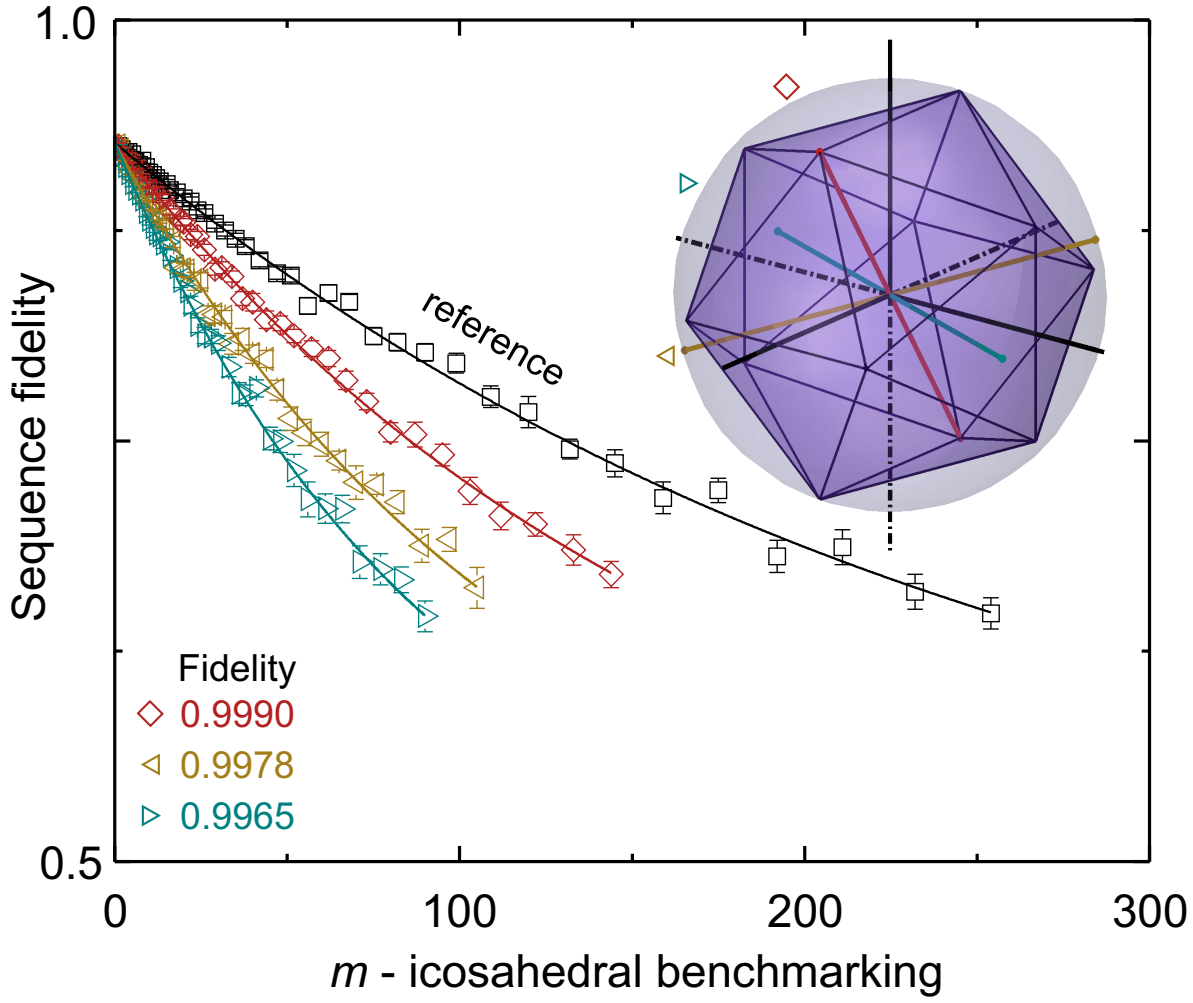


Figure 6.4: (color online) Icosahedral-based randomized benchmarking. We have interleaved three non-Clifford gates whose axes are shown in the inset, the gates rotate around a face center, vertex or edge midpoint of the icosahedron (superimposed). The gates are composed of three, six, and eight elements. Their compositions are: $\mathrm{Y}_{\phi}\ \mathrm{X}_{2\pi/5}\ \mathrm{Y}_{-\phi}$ ($\diamond$), $\mathrm{X}_{\phi}\ \mathrm{Z}_{-2\pi/5}\ \mathrm{Y}\ \mathrm{X}_{2\phi}\ \mathrm{Z}_{2\pi/5}\ \mathrm{X}_{-\phi}$ ($\triangleleft$), and $\mathrm{X}_{\phi}\ \mathrm{Z}_{-2\pi/5}\ \mathrm{X}_{-\phi}\ \mathrm{X}_{-\pi/2}\ \mathrm{Y}_{-\pi/2}\ \mathrm{X}_{\phi}\ \mathrm{Z}_{2\pi/5}\ \mathrm{X}_{-\phi}$ ($\triangleright$). The gate fidelities are tabulated in the figure. The average error per physical gate which makes up the interleaved gates is $r = 3 - 4 \cdot 10^{-4}$.

The rotational axes are highlighted in the inset. The fidelities of these gates are tabulated in the figure. These complex gates work surprisingly well: we compute the average error per physical decomposition to range between $3 \cdot 10^{-4}$ and $4 \cdot 10^{-4}$, assuming that errors are small and uncorrelated. These results demonstrate that even these complex, composite gates can be implemented with high fidelity.

Apart from the first demonstrated implementation of rotational groups beyond the Clifford group, the results on icosahedral benchmarking in Figs. 6.3 and 6.4 clearly indicate that physical rotations, other than the widely used Clifford rotations, can be done with a very similar fidelity. This strongly suggests that any arbitrary rotation can be done with high fidelity. Moreover, the gate parameters can be optimized to achieve decoherence-limited performance using the method outlined in Ref. [**?**], providing an interpolation table for implementing any desired rotation directly, efficiently, and accurately. In addition, icosahedral benchmarking could also be used for evaluating functions of higher order, beyond gate fidelity, as the tetra-, octa-, and icosahedral rotational groups are unitary 2-, 3-, and 5-designs [**? ?**].

6.7 Conclusion

We have shown a quantum version of rolling dice with a superconducting qubit, using gate sets inspired by the Platonic solids. Fundamentally, our work illustrates the potential of using unitaries with a finer distribution for accurate control, and provides a route for the implementation and benchmarking of non-Clifford gates. More generally, our results imply that arbitrary rotations can be done with high accuracy, allowing for complex gates and algorithms to be

performed more efficiently in quantum information processing.

Chapter 7

Overcoming correlated noise in quantum systems: How mediocre clocks make good qubits [1]

7.1 Introduction

One of the main challenges in quantum information is maintaining precise control over the phase. Long-term phase stability therefore seems to be a critical requirement of qubits: good qubits must be "good clocks" [**?**]. Non-Markovian noise, which arises naturally in many quantum systems [**?** **?**], causes frequency drifts which are especially detrimental to long-term stability. Fortunately, this correlated noise can be suppressed using Hahn spin echo [**?**

[1] The following was submitted as: "Overcoming correlated noise in quantum systems: How mediocre clocks make good qubits" Peter O'Malley, Julian Kelly, Rami Barends, et al. (2014).

]. In practice however, optimally suppressing noise in a complex algorithm is limited by the understanding of the noise process. Thus, characterizing noise which acts on short timescales is crucial for understanding the performance of quantum gates, but such noise is difficult to measure precisely with traditional methods such as Ramsey and Hahn spin echo experiments [**? ? ? ?**].

Here, we introduce a metrological tool based on randomized benchmarking (RB) [**? ? ? ?**] to quantify noise on the timescales relevant for quantum gates. We apply it on a superconducting quantum interference device- (SQUID-)based qubit, and show that $1/f$ flux noise [**? ?**] is not a limiting factor, while finding a previously unreported telegraph noise process. We also show that undesired coherent interactions can be understood as an effective correlated noise. Finally, we demonstrate how the effects on gate fidelity of non-Markovian noise can be overcome. The presented approach elucidates how high fidelity quantum operations can be achieved with "mediocre clocks".

Quantum systems based on ion traps, spin qubits, and superconducting circuits are rapidly maturing, with individual operation fidelity at the levels required for fault-tolerant quantum computing [**? ? ? ? ? ? ? ? ?**]. These systems are limited primarily by environmental phase noise, which can manifest as jittering of the qubit's frequency. Noise in the phase ϕ is characterized by variance $\langle \phi^2(\tau) \rangle$, increasing linearly with time τ for white noise, and with higher power for correlated noise [**?**]. Ramsey and spin echo experiments measure the decay of phase coherence over long timescales; at short timescales, dephasing errors are small and thus hard to measure, making physical mechanisms difficult to identify. Here, we quantify

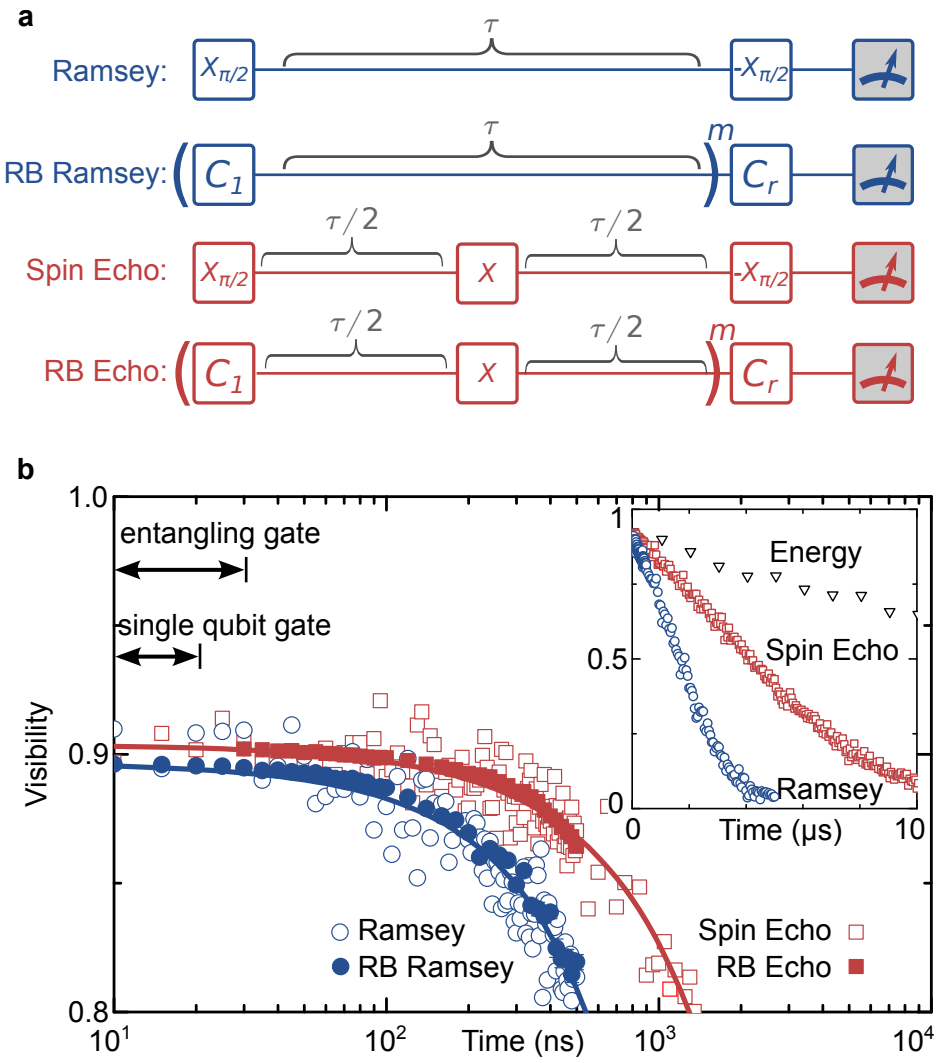


Figure 7.1: (color online) (a) Gate diagram for Ramsey and Hahn spin echo sequences, and their RB equivalents. For RB Ramsey, instead of inserting an idle between $X_{\pi/2}$ pulses, we interleave the idle between Cliffords, after which the qubit is rotated back to the pole and measured. For spin echo and RB echo, an X gate is inserted at the center of the idle. (b) (inset) T_1 (energy decay), Ramsey, and spin echo envelopes. (main) Ramsey (open circle) and spin echo (open square) envelopes at short times. RB decay envelopes have been inferred from $\langle\phi^2(\tau)\rangle$ measured by RB Ramsey (solid circle) and RB echo (solid square); see text for details. The single qubit and entangling gate durations are shown for reference. Note the significantly lower noise of the RB sequences, which take approximately the same measurement time as the Ramsey and echo experiments.

phase noise by using RB to measure the decoherence of an identity gate versus its duration, providing an unprecedented level of precision.

We use a superconducting quantum system based on the planar transmon qubit variant, the Xmon [**?**]. This qubit consists of a SQUID, which serves as a tunable non-linear inductor, and a large X-shaped shunt capacitor. It is well-suited for characterizing phase noise as the qubit has long energy relaxation times, and the presence of a SQUID gives a controllable susceptibility to flux noise. These qubits have frequencies that can be tuned to 6 GHz and below and have typical nonlinearities of $\eta/2\pi = -0.22$ GHz, and capacitive coupling strengths between qubits of $2g/2\pi = 30$ MHz [**?**]. Single qubit rotations are performed with microwave pulses and tuned with ORBIT [**?**], and we use a dispersive readout scheme for state measurements [**?**].

7.2 Randomized benchmarking as a measure of dephasing

Figure 7.1a shows the gate sequences for Ramsey and Hahn spin echo sequences, as well as their RB equivalents, that we have called "RB Ramsey" and "RB echo". The Ramsey experiment accumulates phase error from a single period τ, whereas the RB Ramsey experiment accumulates phase error from m applications of τ, with m typically of order 100. In RB, gate error is measured directly by interleaving gates with random Cliffords, which depolarize errors, such that repeated gate applications add error incoherently [**?**]. Thus, RB Ramsey has a factor m higher sensitivity than Ramsey when errors and times τ are small. The error of an idle gate, $r_{I(\tau)}$, is directly related to the variance of the phase noise by (Appendix D)

$$r_{I(\tau)} = \frac{1}{6}\langle\phi^2(\tau)\rangle. \tag{7.1}$$

Using this relation, we infer and plot the equivalent Ramsey decay envelope visibility V (solid circles) with $V(\tau) = A\exp(-1/2\langle\phi^2(\tau)\rangle) + B$ in Fig. 7.1b, with preparation and measurement error parameters A and B extracted from the Ramsey fit. We likewise show the equivalent spin echo decay envelope from RB echo data as solid squares. The Ramsey and spin echo measurements over the same timescale are shown for comparison as open circles and open squares, respectively. We label the length of a single qubit and two-qubit entangling gate [**?**] to emphasize the relevant timescale. The full Ramsey and spin echo measurements are shown on the typical linear scale, together with energy relaxation, in the inset of Fig. 7.1b.

As shown in Fig. 7.1b, the RB Ramsey and RB echo data are consistent with the Ramsey and spin echo measurements, respectively, while being much more precise. Any structure to short-time dephasing is obscured in the Ramsey data, whereas the RB Ramsey data reveal a time dependence that we will show is consistent with telegraph noise. The use of RB greatly improves the precision of phase noise measurements; the range of the measured Ramsey visibility for $\tau < 300$ ns is reduced by an order of magnitude. We note that the amount of total time taken to perform the Ramsey and RB Ramsey measurements is approximately the same.

7.3 Identifying telegraph noise

In order to identify the dominant noise mechanism, we examine the dependence of idle gate error on time and compare against different noise models in Fig. 7.2. Whereas in Fig. 7.1 we infer an equivalent Ramsey envelope, here we plot the idle gate error directly, as measured by RB Ramsey (T_1 effects have been subtracted) (Appendix D). For short times, we see a non-

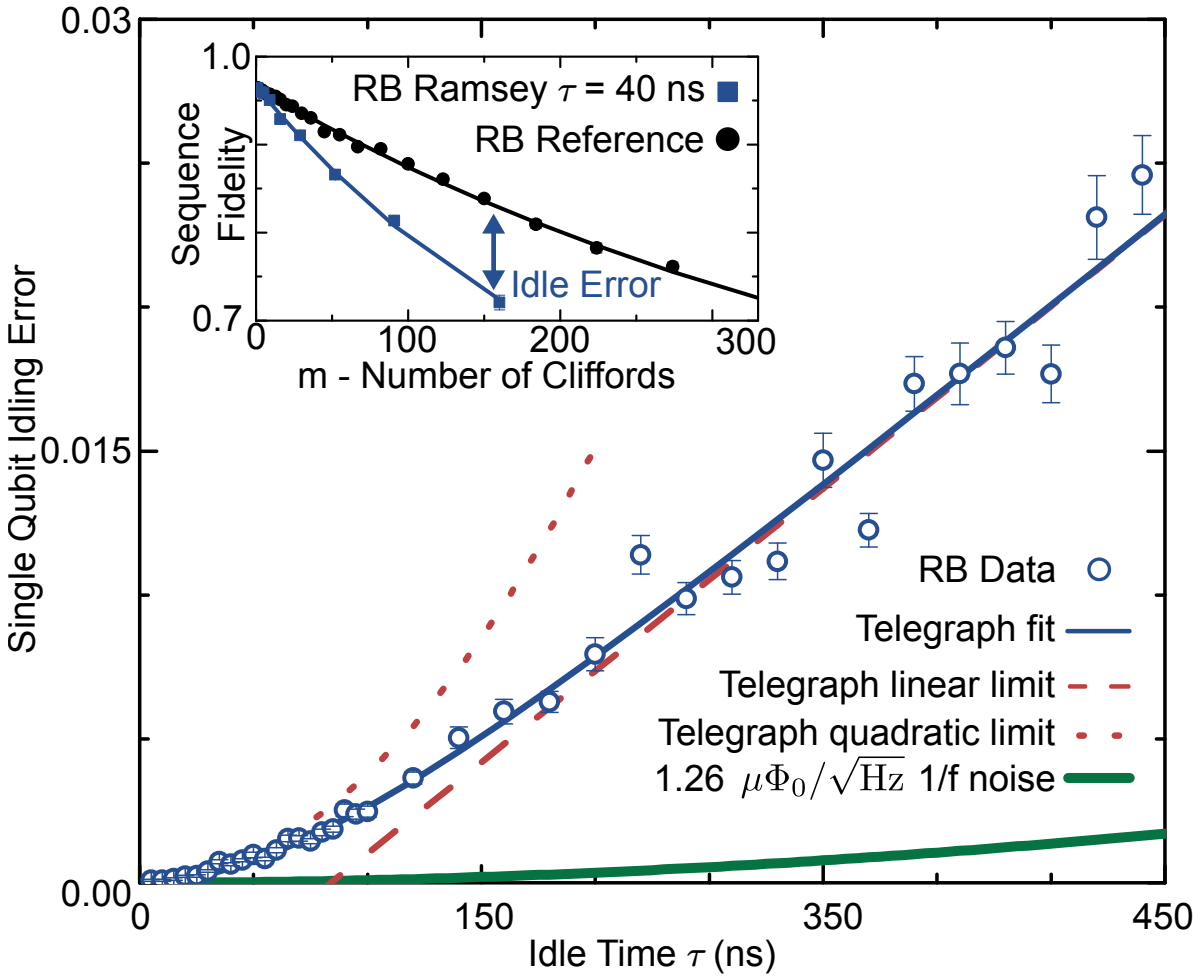


Figure 7.2: (color online) RB Ramsey measurement (circles) for short timescales; note that the small error from T_1 decay, which is 9×10^{-4} at 450 ns, has been subtracted (Appendix D). We fit to a telegraph noise model (Eq. 7.4); the dotted (dashed) lines give the short (long) time limit of the noise model. The inferred but negligible contribution from $1/f$ noise as measured for this qubit (Appendix D) is shown as a thick line. The inset shows the experiment used to extract the 40 ns data point.

linear increase of error with gate duration which transitions into a linear behavior for lengths above approximately 100 ns. The inset shows the sequence fidelity *vs.* number of Cliffords, with and without interleaved idles, used to extract the idle error for $\tau = 40$ ns.

While it has long been known that SQUIDs are susceptible to $1/f$ flux noise [**? ? ? ? ? ? ?**], we find that this has a negligible contribution to gate error. A system limited by $1/f$ and white noise would see a linear increase in error at short, and quadratic increase at long times, as the $1/f$ component begins to dominate. The data exibit the opposite trend. Moreover, the expected contribution to gate error from $1/f$ noise, as measured for this system below 1 Hz using the Ramsey Tomography Oscilloscope protocol [**?**] (Appendix D), is significantly less than observed here (Fig. 7.2 thick solid line).

The trend observed in Fig. 7.2 is consistent with telegraph noise. For a random telegraph switching of the qubit frequency, the phase noise is given by

$$\langle \phi^2_{\text{tel}}(\tau) \rangle = 2(2\pi\Delta f_{10})^2 T_{\text{sw}} \left(\tau - T_{\text{sw}} \left[1 - \exp\left(-\frac{\tau}{T_{\text{sw}}} \right) \right] \right), \tag{7.2}$$

where Δf_{10} is the effective switching amplitude of the qubit frequency and T_{sw} is the switching timescale (Appendix D). In a more general case, the error rate for an idle of length τ, $r_{I(\tau)}$, can be fit to a combination of error sources: white, long-time correlated, $1/f$, and telegraph phase noise, as well as T_1 decay,

$$r_{I(\tau)} = \frac{\tau}{3T_1} + \frac{1}{6}\Big(\langle \phi^2_{\text{white}}(\tau) \rangle + \langle \phi^2_{\text{corr}}(\tau) \rangle + \langle \phi^2_{\text{1/f}}(\tau) \rangle + \langle \phi^2_{\text{tel}}(\tau) \rangle \Big), \tag{7.3}$$

where the derivation for $\langle\phi^2_{\text{white}}(\tau)\rangle = 2\tau/T_{\phi 1}$, $\langle\phi^2_{\text{corr}}(\tau)\rangle = 2(\tau/T_{\phi 2})^2$, and $\langle\phi^2_{\text{1/f}}(\tau)\rangle$ are given in(Appendix D), and we assume correlated noise has a longer timescale than the experiment. The data here are fitted to a noise model featuring only T_1 decay and telegraph noise,

$$r_{I(\tau)} = \frac{\tau}{3T_1} + \frac{1}{6}\langle\phi^2_{\text{tel}}(\tau)\rangle. \tag{7.4}$$

We extract $T_{\text{sw}} = 84 \pm 14$ ns, and $\Delta f_{10} = 479 \pm 30$ kHz from the fit. The dotted (dashed) line shows this noise model in the short (long) time limit. Perhaps surprisingly, this measurement directly shows that gates of duration 20 ns can achieve fidelity > 0.999 in a system with characteristic Ramsey scale of $T_{\phi 2} = 2.0\,\mu$s (Appendix D).

One potential source of telegraph frequency fluctuations in transmon qubits is charge noise [**?** **?**]. Charge noise is exponentially suppressed when $E_J/E_C \gg 1$; for our qubit, the calculated frequency splitting ranges from 1 Hz to 14 kHz (Appendix D), well below the magnitude necessary to explain the data. Additionally, we observe a general trend of increasing noise with decreasing qubit frequency, which points away from resonator-induced dephasing being a dominant source of noise (Appendix D). As the data are entirely fit by a telegraph noise model, this suggests a new mechanism is at work.

7.4 Coherent dephasing

Coherent errors arising from unwanted qubit-qubit interactions can also contribute to dephasing [**?**]. In Fig. 7.3, we explore these effects in our system. Figure 7.3a shows an energy level diagram for capacitively coupled qubits, where the fundamental entangling rate Ω_{ZZ} [**?**] arises from an avoided level crossing between the $|11\rangle$ state and the $|02\rangle$ and $|20\rangle$ states. This

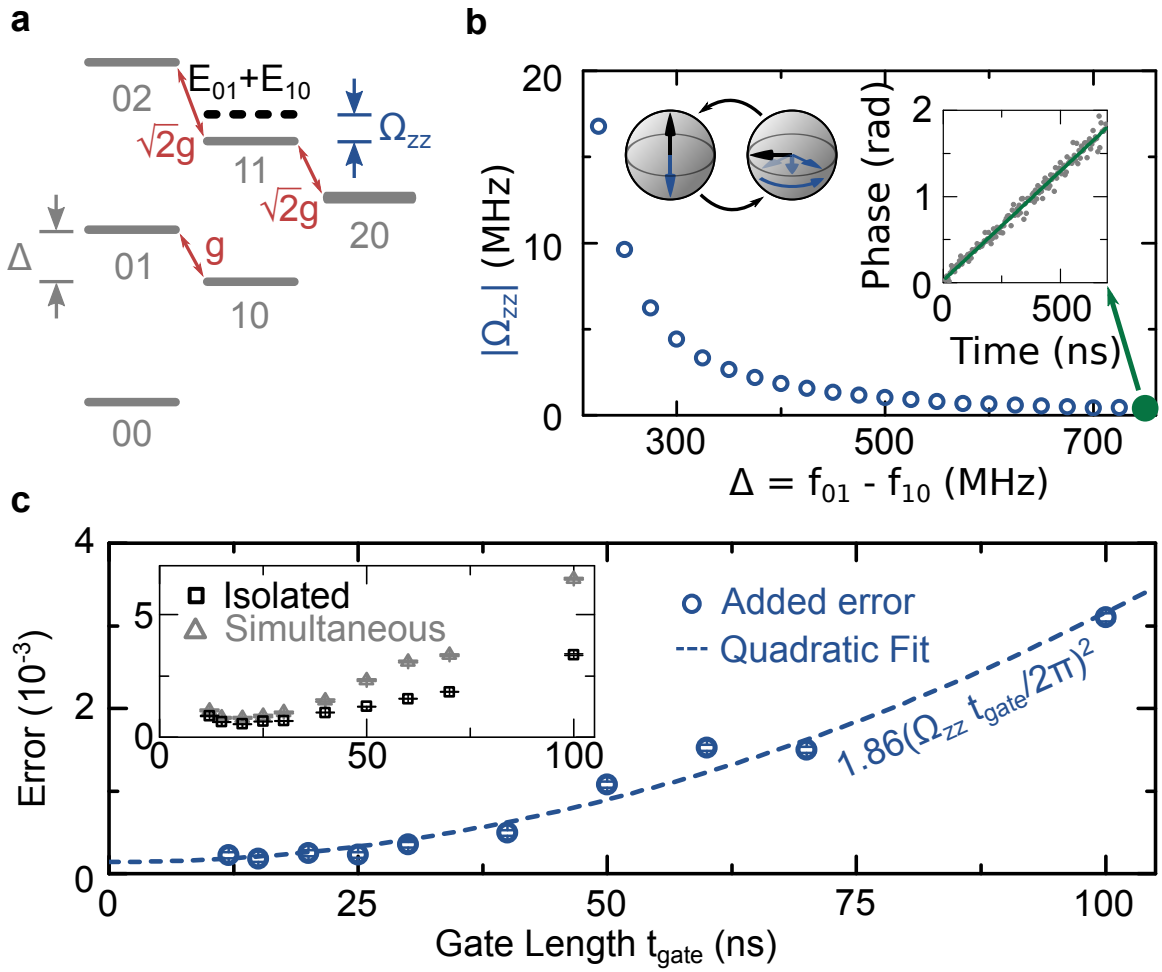


Figure 7.3: (color online) (a) Energy level diagram for two capacitively coupled qubits with coupling strength $2g/2\pi = 30$ MHz, detuned by frequency Δ. The avoided level crossing between the $|11\rangle$ and $|02\rangle/|20\rangle$ states repels the $|11\rangle$ frequency from the sum of $|01\rangle$ and $|10\rangle$ frequences by the amount Ω_{ZZ}. (b) This entangling interaction causes the phase of one qubit to precess, conditional on the state of its neighbor (cartoon and inset). The Ω_{ZZ} interaction decreases with Δ, to a level of $\Omega_{ZZ}/2\pi = 0.4$ MHz at $\Delta/2\pi = 750$ MHz. (c) RB data isolating the Ω_{ZZ} interaction. Gate error is measured *vs.* gate duration for a single qubit and when qubits are operated simultaneously (inset). The difference (main figure) measures the error contribution from the Ω_{ZZ} interaction, and is fit to $1.86(\Omega_{ZZ}t_{\text{gate}}/2\pi)^2 + 1.4 \cdot 10^{-4}$.

interaction manifests as a state-dependent frequency shift, that falls off with detuning Δ, as measured in Fig. 7.3b. We note that in the case of a qubit coupled to a resonator, Ω_{ZZ} is equivalent to the dispersive shift [**?**] 2χ as defined in [**?**]. The inability to turn this interaction off completely results in additional errors when operating qubits simultaneously. Figure 7.3c shows average gate error *vs.* duration, when a qubit is operated in isolation or simultaneously with a coupled qubit ($\Omega_{ZZ}/2\pi$ = 0.4 MHz). Error for single qubit or simultaneous operation is inferred from the RB reference error per Clifford, divided by the average of 1.875 physical gates per Clifford [**?**]. The difference between isolated and simultaneous operation gives the added error from the Ω_{ZZ} interaction, which is fit to a quadratic.

This interaction is correlated, and therefore the errors are quadratic with gate duration; specifically, the error per gate due to the Ω_{ZZ} interaction between two qubits simultaneously undergoing RB is

$$E = \frac{\pi^2}{6}\left(\frac{\Omega_{ZZ}}{2\pi}t_{\text{gate}}\right)^2, \tag{7.5}$$

where $\Omega_{ZZ}/2\pi$ is the magnitude of the interaction and t_{gate} is the RB gate duration (Appendix D). The fit to the data has a quadratic coefficient of 1.86 ± 0.1, while $\pi^2/6 \approx 1.64$. Here, the careful application of RB both distinguishes these errors at the $1 \cdot 10^{-4}$ level, and indicates that short gates are effective in suppressing them.

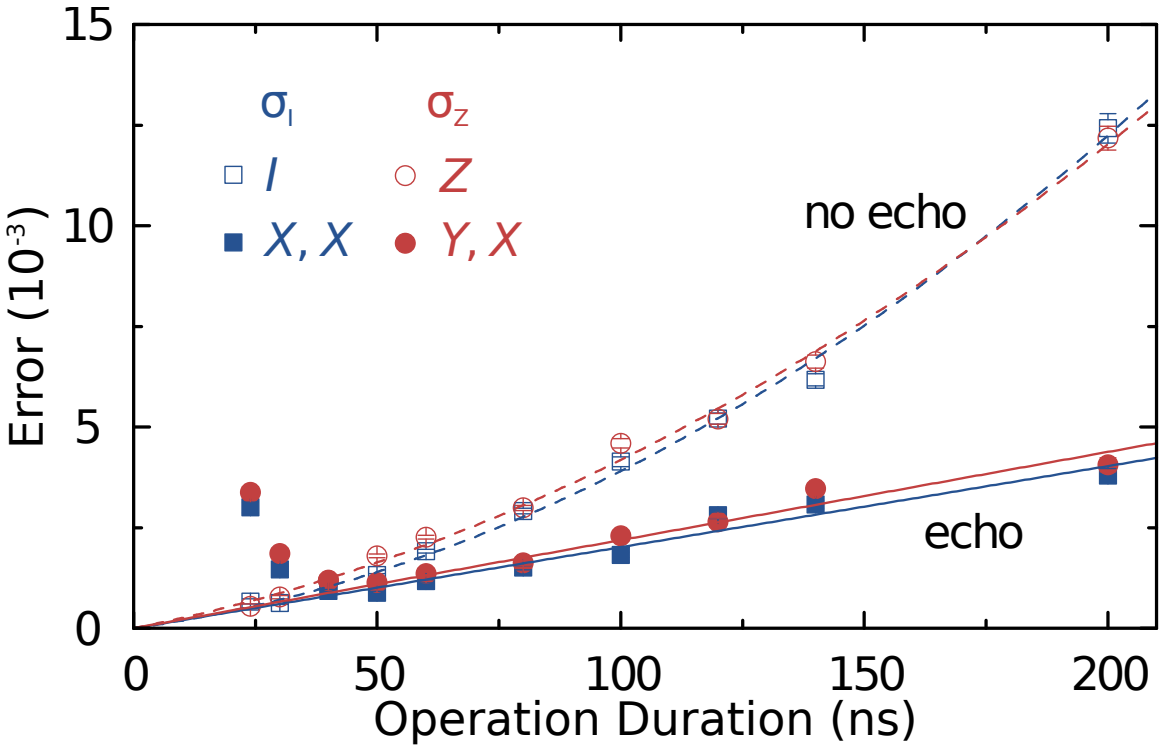


Figure 7.4: (color online) Operation error of σ_I and σ_Z, implemented with (closed symbols) and without (open symbols) echoing, as measured with interleaved RB. The data are fitted to a linear and quadratic form, representing uncorrelated and correlated noise. Short gates and echoing allow for reaching low error, limited by uncorrelated noise.

7.5 Optimal gates

We now examine the behavior of gate fidelity for a variety of gates in the presence of the non-Markovian noise we have measured. Fig. 7.4 shows gate fidelity *vs.* gate length for two implementations each of two different gates: for σ_I, an idle and two microwave pulses (X, X), and for σ_Z, a frequency detuning pulse and two microwave pulses (Y, X). The errors of these operations *vs.* duration are determined with interleaved RB. In agreement with previous measurements, we find that the error of operations without X or Y pulses (open symbols) follow a quadratic-like dependence with gate duration at these timescales. Using X or Y pulses (closed symbols), we observe a linear-like dependence at longer durations, indicating that the correlated part of the phase noise has been suppressed. Below 40 ns, we find an increased error which we attribute to the population of higher levels due to spectral leakage [**?**]. The solid (dashed) lines are linear (linear and quadratic) fits to the data.

We find that T_1 effects contribute approximately half of the linear error accumulation seen in Fig. 7.4, with the remainder being equivalent to a white noise dephasing timescale of $T_{\phi 1} \approx 30\,\mu$s, while the quadratic terms present in the I and Z cases correspond to a correlated dephasing timescale of $T_{\phi 2} \approx 1\,\mu$s (Appendix D). These timescales describe the gate behavior much more accurately than those given by Ramsey and spin echo measurements (Appendix D), again indicating that RB can measure effects that Ramsey and spin echo are insensitive to. With the accurate characterization enabled by RB, we find that either short gates, or long gates with intrinsic echoing, are effective at remedying non-Markovian noise.

7.6 Conclusion

We have taken RB, a protocol for determining the fidelity of gates, and applied it as a metrological tool for identifying noise processes. Applied to a superconducting qubit system, we have found a previously unreported telegraph noise mechanism, accurately measured dephasing caused by coherent qubit-qubit interactions, and determined the optimal implementation of different quantum gates. Our results show that non-Markovian phase noise can be effectively suppressed through short gates and echoing, allowing mediocre clocks to perform high-fidelity gates.

Chapter 8

State preservation by repetitive error detection in a superconducting quantum circuit[1]

8.1 Introduction

Quantum computing becomes viable when a quantum state can be preserved from environmentally-induced error. If quantum bits (qubits) are sufficiently reliable, errors are sparse and quantum error correction (QEC) [**?** **?** **?** **?** **?** **?**] is capable of identifying and correcting them. Adding more qubits improves the preservation by guaranteeing increasingly larger clusters of errors will not cause logical failure – a key requirement for large-scale systems. Using QEC

[1]The following was published as: "State preservation by repetitive error detection in a superconducting quantum circuit" Julian Kelly, Rami Barends, Austin Fowler, et al. *Nature* 519, 66-69 (2014).

to extend the qubit lifetime remains one of the outstanding experimental challenges in quantum computing. Here, we report the protection of classical states from environmental bit-flip errors and demonstrate the suppression of these errors with increasing system size. We use a linear array of nine qubits, which is a natural precursor of the two-dimensional surface code QEC scheme [**?**], and track errors as they occur by repeatedly performing projective quantum non-demolition (QND) parity measurements. Relative to a single physical qubit, we reduce the failure rate in retrieving an input state by a factor of 2.7 for five qubits and a factor of 8.5 for nine qubits after eight cycles. Additionally, we tomographically verify preservation of the non-classical Greenberger-Horne-Zeilinger (GHZ) state. The successful suppression of environmentally-induced errors strongly motivates further research into the many exciting challenges associated with building a large-scale superconducting quantum computer.

The ability to withstand multiple errors during computation is a critical aspect of error correction. We define n-th order fault-tolerance to mean that *any* combination of n errors is tolerable. Previous experiments based on nuclear magnetic resonance [**?** **?**], ion traps [**?**], and superconducting circuits [**?** **?** **?**] have demonstrated multi-qubit states that are first-order tolerant to one type of error. Recently, experiments with ion traps and superconducting circuits have shown the simultaneous detection of multiple types of errors [**?** **?**]. The above hallmark experiments demonstrate error correction in a single round; however, quantum information must be preserved throughout computation using multiple error correction cycles. The basics of repeating cycles have been shown in ion traps [**?**] and superconducting circuits [**?**]. Until now, it has been an open challenge to combine these elements to make the information stored

in a quantum system robust against errors which intrinsically arise from the environment.

The key to detecting errors in quantum information is to perform QND parity measurements. In the surface code, this is done by arranging qubits in a chequerboard pattern – with data qubits corresponding to the white, and measure qubits to the black squares (see Fig. 8.1) – and using these ancilla measure qubits to repetitively perform parity measurements to detect bit-flip ($\hat{X}$) and phase-flip ($\hat{Z}$) errors [**?**]. A square chequerboard with $(4n+1)^2$ qubits is n-th order fault tolerant, meaning at least $n+1$ errors must occur to cause failure in preserving a state if fidelities are above a threshold. With error suppression factor $\Lambda > 1$ and more qubits, failure becomes increasingly unlikely with probability $\epsilon \sim 1/\Lambda^{n+1}$ (assuming independent errors). The surface code is highly appealing for superconducting quantum circuits as it requires only nearest neighbour interactions, single and two-qubit gates, and fast repetitive measurements with fidelities above a lenient threshold of approximately 99%. All of this has recently been demonstrated in separate experiments [**? ?**].

8.2 Experiment design

With the results of chapter 4 combined with the technology developed in Ref. [**?**], all of the elements required to implement the multi-qubit $\hat{Z}\hat{Z}$ or $\hat{X}\hat{X}$ operators outlined in chapter 9 have been realized, making it possible to perform experiments related to the surface code. At this point, one might think that the most natural step is to create a two-dimensional array of qubits, as in the surface code, to build a fault-tolerant logical qubit. This is challenging for two reasons. First, fabricating the wiring necessary to implement a two-dimensional array of qubits

is non-trivial, as scalable wiring system requires routing wires out of the plane that contains the qubits. This can be easily understood as routing the wires to the perimeter will not work for large numbers of qubits, given the scaling of the length of the perimeter of a square as compared to the area. This technology, forecasted to be a complex but feasible engineering challenge, does not yet exist. Secondly, implementing a full logical qubit with both $\hat{X}$ and $\hat{Z}$ protection is a sophisticated experiment, unsuited for a first attempt. Instead, we test the basics of the surface code in a simpler experiment.

8.3 Repetition code: device and algorithm

The simplest system demonstrating the basic elements of the surface code is a one-dimensional chain of qubits, as seen in Fig. 8.1a. It can run the repetition code, a primitive of the surface code, which corrects bit-flip errors on both data and measure qubits. The device shown in Fig. 8.1b is a chain of nine qubits, which allows us to experimentally test both first- and second-order fault-tolerance. It consists of a superconducting aluminium film on a sapphire substrate, patterned into Xmon transmon qubits [**?**] with individual control and readout. The qubits are the cross-shaped devices; they are capacitively coupled to their nearest neighbours, controlled with microwave drive and frequency detuning pulses, and measured with a dispersive readout scheme. The device consists of five data qubits and four measure qubits in an alternating pattern, see Appendix E for details.

To detect bit-flips, we determine the parity of adjacent data qubits by measuring the operator $\hat{Z}\hat{Z}$. We do this using an ancilla measure qubit, and performing single- and two-qubit quantum

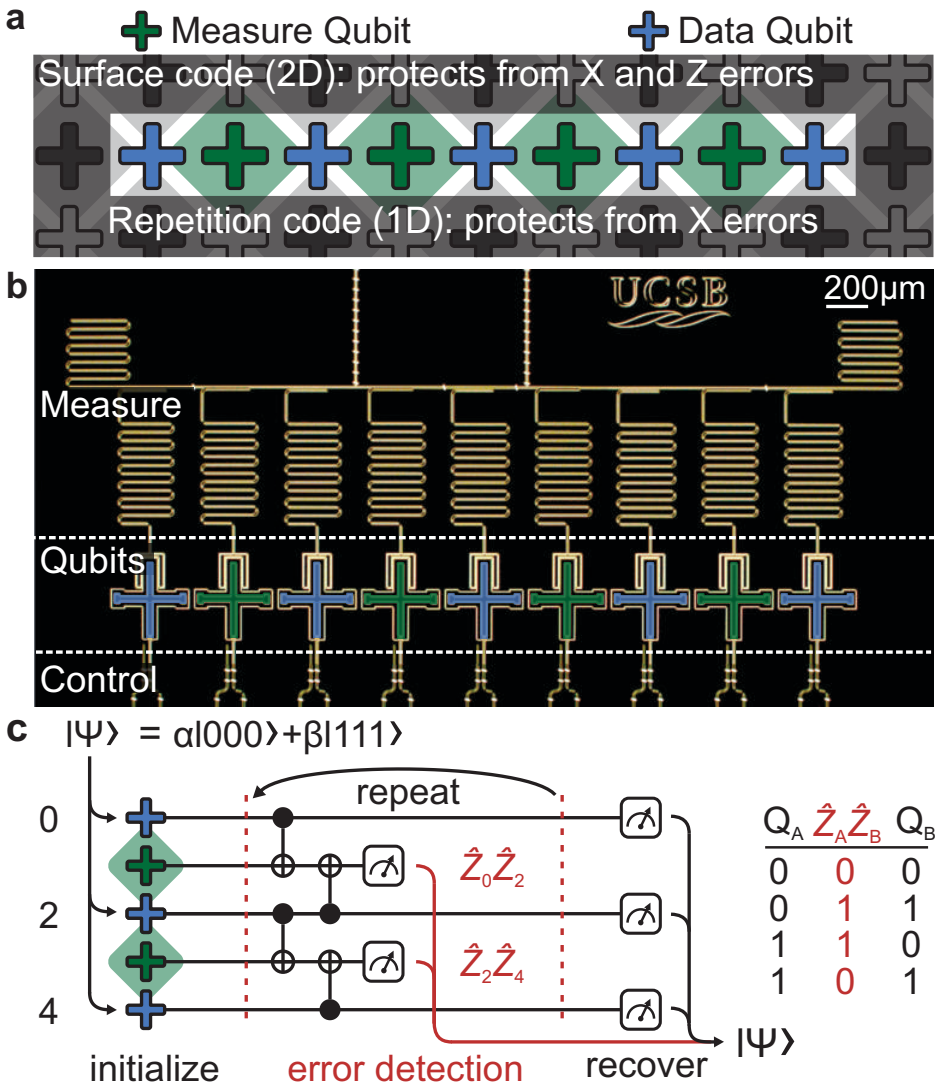


Figure 8.1: **Repetition code: device and algorithm.** (a) The repetition code is a one-dimensional variant of the surface code, and is able to protect against $\hat{X}$ (bit-flip) errors. The code is implemented using an alternating pattern of data and measure qubits. (b) Optical micrograph of the superconducting quantum device, consisting of nine Xmon [**?**] transmon qubits with individual control and readout, with a nearest-neighbour coupling scheme. (c) The repetition code algorithm uses repeated entangling and measurement operations which detect bit-flips, using the parity scheme on the right. Using the output from the measure qubits during the repetition code, the initial state can be protected by detecting physical errors. Measure qubits are initialized into the $|0\rangle$ state and need no reinitialization as measurement is QND.

gates (Fig. 8.1c). The operator measurement has the value -1 and leaves the ancilla qubit state unperturbed for states $|00\rangle$ and $|11\rangle$, and value +1 which flips the ancilla qubit state for $|01\rangle$ and $|10\rangle$. Therefore, errors can be detected as they occur by repeating this operator and noting changes in the outcome. Importantly, this measurement does not destroy the quantum nature: given input $\alpha|00\rangle + \beta|11\rangle$ the result is -1 and the quantum state remains, with similar behavior for other Bell-like superposition states. In the repetition code, simultaneous measurements of these operators enable multiple bit-flip errors to be detected.

8.4 Identifying bit-flip errors

We now discuss how bit-flip errors, which can occur on any qubit and at any time, are identified. The quantum circuit of the repetition code is shown in Fig. 8.2a, for three cycles (in time) and nine qubits. This is the natural extension of the schematic in Fig. 8.1c, optimized for our hardware (Appendix E). The figure illustrates four distinct types of bit-flip errors (stars): measurement error (gold), single-cycle data error (purple), two-cycle data error (red), and a data error after the final cycle (blue). Controlled-NOT (CNOT) gates propagate bit-flip errors on the data qubit to the measure qubit. Each of these errors is typically detected at two locations if in the interior and one location if at the boundary; we call these "detection events". The error connectivity graph [**?**] is shown in Fig. 8.2b, where the grey lines indicate every possible pattern of detection events that can arise from a single error. The last column of values for the $\hat{Z}\hat{Z}$ operators in Fig. 8.2b are constructed from the data qubit measurements, so that errors between the last cycle and data qubit measurement can be detected (Appendix E).

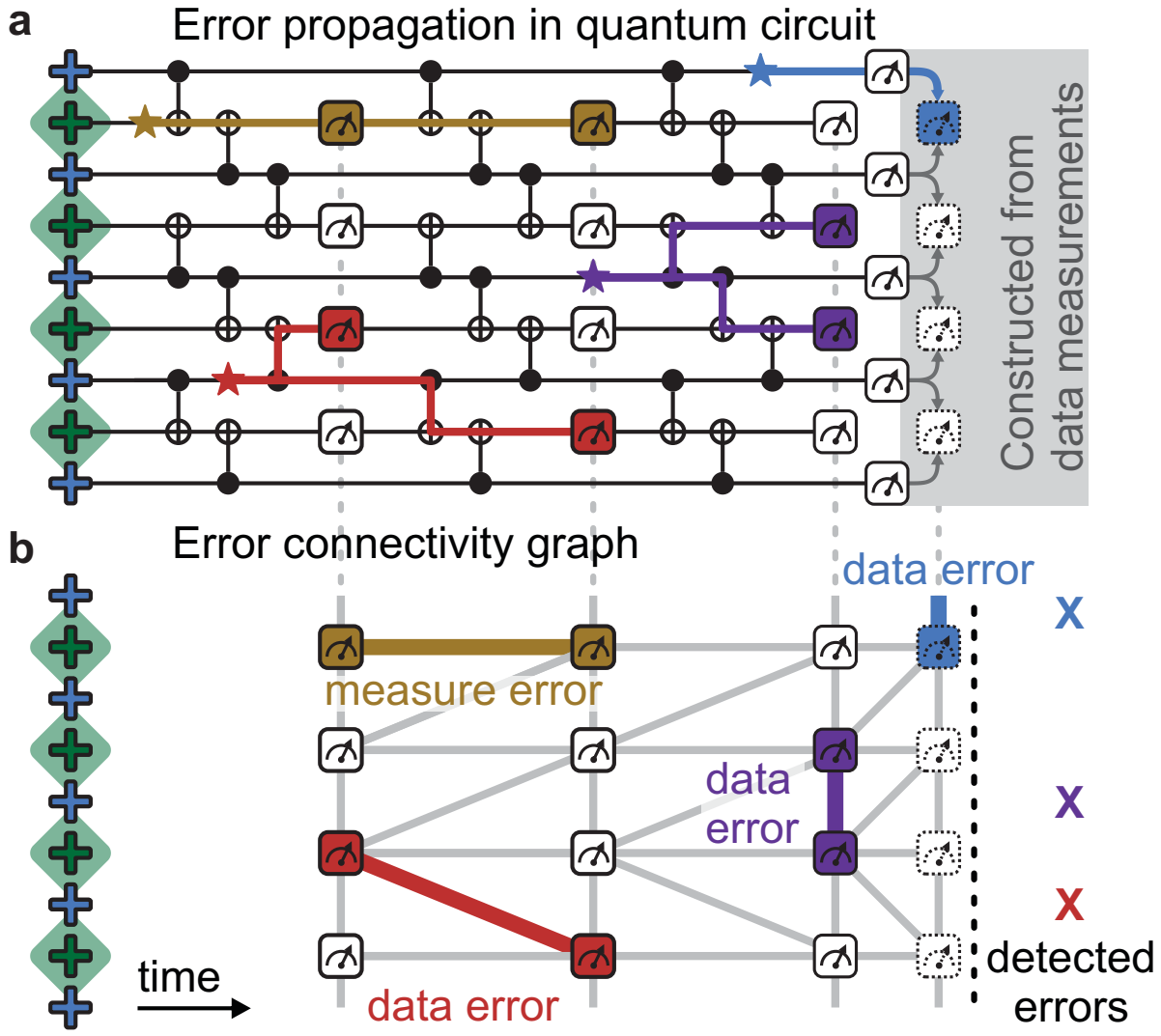


Figure 8.2: **Error propagation and identification.** (a) The quantum circuit for three cycles of the repetition code, and examples of errors. Errors propagate horizontally in time, and vertically through entangling gates. Different errors lead to different detection patterns: An error on a measure qubit (gold) is detected in two subsequent rounds. Data qubit errors (purple, red, blue) are detected on neighbouring measurement qubits in the same or next cycle. Data errors after the last round (blue) are detected by constructing the final set of $\hat{Z}\hat{Z}$ eigenvalues from the data qubit measurements. (b) The connectivity graph for the quantum circuit above, showing measurements and possible patterns of detection events (grey), see text. The example detection events and their connections are highlighted, the corresponding detected errors are shown on the right, which when applied, will recover the input data qubit state.

In the absence of errors, there are two possible patterns of sequential measurement results. If a measure qubit's neighbouring data qubits are in the $|00\rangle$ or $|11\rangle$ state, the measure qubit will report a string of identical values. If the data qubits are in the $|01\rangle$ or $|10\rangle$ state, the measure qubit will report alternating values, as measurement is QND. Single data bit-flip errors make the measurement outcomes switch between these two patterns. For example, if the measurement outcomes for three cycles are 0, 0, and 1, this indicates a change from the identical to the alternating pattern in the last measurement, and hence a detection event. Explicitly, with m_t the measure qubit outcome at cycle t and $\oplus$ the exclusive OR (XOR) operator, for each of the two patterns we have $b_t = m_{t-1} \oplus m_t = 0$ or 1. A detection event at cycle t is then identified when $D_t = b_{t-1} \oplus b_t = 1$.

We use minimum-weight perfect matching [**? ? ?**] to decode to physical errors, based on the pattern of detection events and an error model for the system. Intuitively, it connects detection events in pairs or to the boundary using the shortest weighted path length. It is important to note that errors can lead to detection event pairs that span multiple cycles, necessitating the need for multi-round analysis as opposed to round-by-round. See Appendix E for details.

8.5 The repetition code and GHZ states

To study the ability of our device to preserve quantum states, we initialised the data qubits into a GHZ state $[(|000\rangle + |111\rangle)/\sqrt{2}]$ and applied two rounds of the repetition code, see Fig. 8.3. The algorithm is shown in Fig. 8.3a. Using quantum state tomography we measured the input GHZ state to have a fidelity $\mathrm{Tr}\,(\rho_{\mathrm{ideal}}\rho)$ of 82%, above the threshold of 50% for

genuine entanglement [**?**]. After two repetition code cycles, we use tomography to construct the density matrices for each pattern of detection events. We find a state fidelity of 78% in the case of no detection events, indicating a retention of genuine quantum entanglement. In the case of two detection events, which indicate a likely data qubit error in the first cycle, we find elements away from the ideal positions. By applying the recovery operation in postprocessing (a single bit-flip on the blue data qubit) we can restore the state. We find that the off-diagonal elements have not vanished – and genuine entanglement is preserved with a fidelity of 59% – even though the repetition code does not provide phase protection. Reduced but non-zero off-diagonal terms indicate bit errors arise from incoherent processes, like qubit energy relaxation which scrambles the phase, as well as coherent processes. Conditional tomography for every configuration can be found in the Appendix E.

The data in Fig. 8.3 clearly show that the one-dimensional repetition code algorithm does not necessarily destroy the quantum nature of the state. It allows for preserving the quantum state in the case of no errors, and correcting bit-flip errors otherwise. This preservation is achieved purely through error detection and classical post-processing, like for the full surface code, avoiding the need for dynamic feedback with quantum gates.

8.6 Repetition code performance

We now address the critical question of how well our implementation of the repetition code protects logical states over many cycles. The process flow is illustrated in Fig. 8.4a. We start by initialising the data qubits in either of the logical basis states: $|0_L\rangle = |0..0\rangle$ or $|1_L\rangle = |1..1\rangle$.

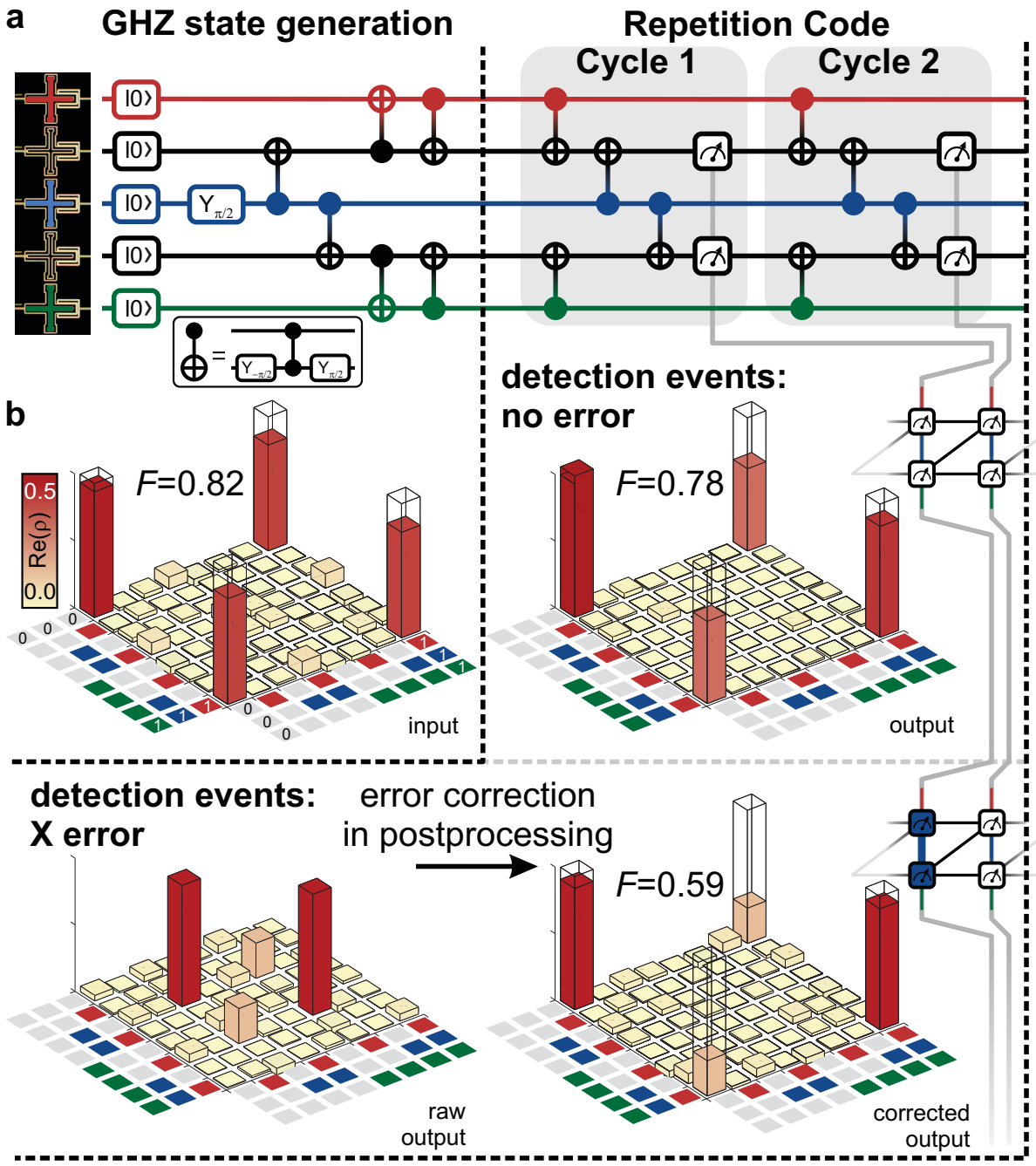


Figure 8.3: **Protecting the GHZ state from bit-flip errors.** (a) Quantum circuit for generating the GHZ state and two cycles of the repetition code. CNOT gates are physically implemented with controlled-phase (CZ) and single qubit gates. (b) Quantum state tomography on the input, and after the repetition code conditional on the detection events: We input a GHZ state with a fidelity of 82%, and find, for the case of no detection events, a 78% fidelity GHZ state. For the detection event connecting both measure qubits, indicating a likely bit-flip error on the central data qubit, we find that through correcting in postprocessing by exchanging matrix elements we recover the major elements of the diagonal. We also recover nonzero off-diagonal elements, indicating some bit-flip errors are coherent.

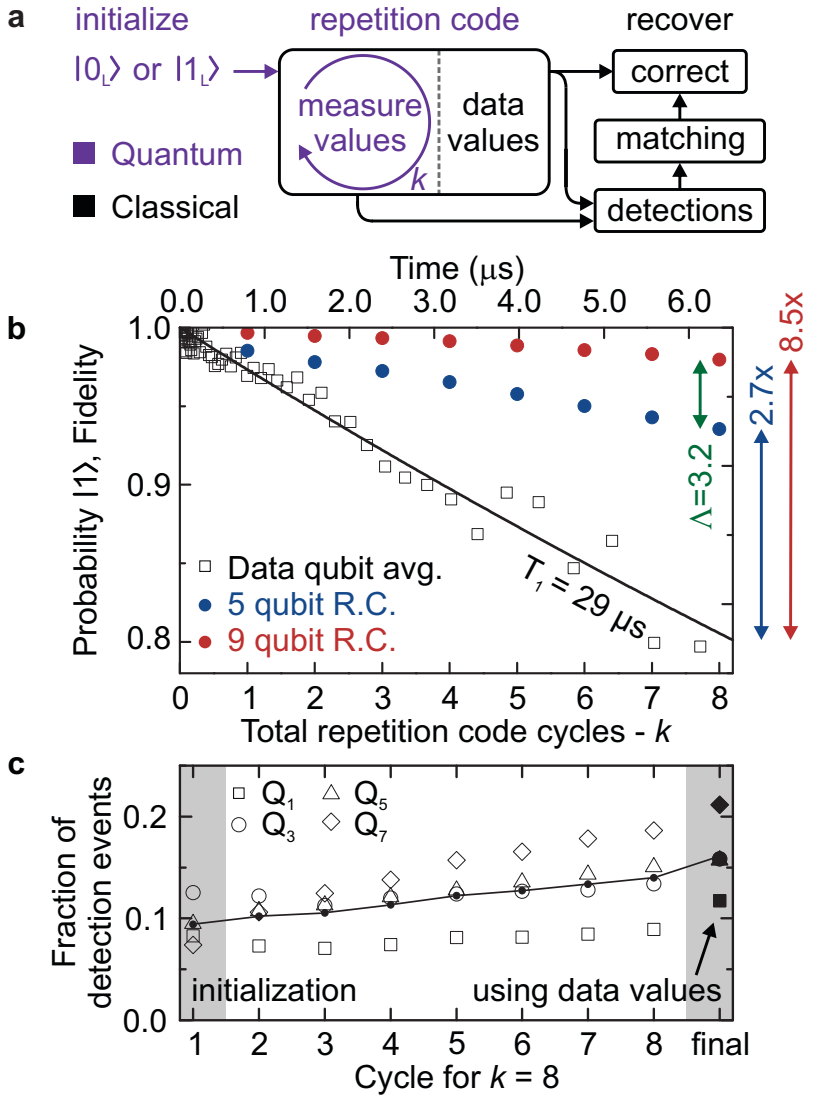


Figure 8.4: **Logical state preservation with the repetition code.** (a) Information flowchart of the repetition code. The data qubits are initialised into $|0_L\rangle$ or $|1_L\rangle$, and the repetition code is repeated k times. In post-processing, the measurement qubit outcomes are converted into detection events and matched to find likely errors, see Fig. 8.2. A successful recovery converts the measured data qubit state into the input state. (b) Memory fidelity $vs.$ time and cycles for a single physical qubit (black) and the five- (blue) and nine- (red) qubit repetition code. Note that energy relaxation decays from a fidelity of 1 to 0, whereas the repetition code decays from a fidelity of 1 to 0.5. Five qubit data sampled from nine qubit data, see Appendix E. The average physical qubit lifetime is $T_1 = 29$ μs, and after eight cycles we see an improvement in error rate by a factor of 2.7 for five qubits, and 8.5 for nine qubits when using the repetition code. This indicates a Λ parameter of 3.2 for our system after eight cycles. (c) Average number of detection events per measure qubit, $vs.$ cycle number, for experiments consisting of eight cycles. We see an increasing rate of detection events with increasing cycle number. This can be attributed to the statistically increasing number of odd parity $\hat{Z}\hat{Z}$ measurements, see text.

We then run the repetition code algorithm for k cycles, and finish by measuring the state of all data qubits. We repeat this 90,000 times to gather statistics. The classical measurement results are converted into detection events, which are processed using minimum-weight perfect matching to generate corrections, see Appendix E. These corrections are then applied to the measured data qubit output to see if the input is recovered. Due to the topological nature of errors, we either recover the logical state, or the bit-wise inverse (see Appendix E). The fidelity of the repetition code algorithm is defined by the success rate of this recovery. In our system, qubits naturally relax to $|0\rangle$, intrinsically making $|0_L\rangle$ more robust than $|1_L\rangle$. To balance these errors and to increase the worst-case lifetime of the system, we apply physical bit-flips to each data qubit at the end of each cycle. This logical flip is compensated for in software. In order to quantify the reduction of logical errors with system size n, we have implemented the repetition code with five and nine qubits in total, corresponding to first- and second-order fault-tolerance.

In Fig. 8.4b we show the fidelity of the repetition code as a function of the number of cycles for five (blue) and nine (red) qubits. We also plot the probability of a $|1\rangle$ state idling for the same duration, averaged over the five data qubits (black). This allows for a direct comparison of single physical qubit error with the multi-qubit logical error. We find a reduced error of logical states after eight cycles as compared to a physical qubit; by a factor of 2.7 for five qubits and 8.5 for nine qubits. We also see a non-exponential fidelity decay for logical states, due to an increasing error rate with cycle number, see Figure 8.4c. We attribute this to energy relaxation of measure qubits. Initial logical states of all 0's or 1's have even parity for all $\hat{Z}\hat{Z}$ operators, maintaining the initial measure qubit $|0\rangle$ state. A bit-flip error on a data qubit,

statistically more likely with increasing cycle number, will cause the nearby $\hat{Z}\hat{Z}$ operators to have odd parity. This will flip measure qubits between the $|0\rangle$ and $|1\rangle$ state at each cycle, making them susceptible to energy relaxation and hence increasing the rate of detection events, see Appendix En.

Figure 8.4 demonstrates state preservation through error correction. We emphasize that we correct errors that intrinsically arise from the environment. Additionally, we see larger repetition codes leading to greater error suppression. This is evidence for the system operating with fidelities above the repetition code threshold. As the error rates grow with cycle number, the many-cycle behaviour of the repetition code must be explored to ensure that the the system remains above threshold. The ratio of the errors for the $n = 1$ and $n = 2$ case after eight cycles suggests $\Lambda = 3.2$, but larger system sizes are needed to infer this accurately for large n and verify that the logical error rate is suppressed exponentially as $\epsilon_{\text{logical}} \sim 1/\Lambda^{n+1}$, as desired.

8.7 Conclusion

Our demonstration that information can be stored with lower error in logical states than in single physical qubits shows that the basic physical processes required to implement surface code error correction are technologically feasible. We hope our work helps accelerate research into the many outstanding challenges that remain, such as development of two-dimensional qubit arrays, improving gate and measurement fidelities [**?**], and investigating the many-cycle behavior of error correction schemes.

Chapter 9

Outlook

In Chapter 9, I outlined some of the challenges which must be overcome for quantum computing to become a practical reality. Due to the analog nature of quantum states, and the existence of decoherence, it was unclear whether we will ever see a quantum computer made. Therefore, it becomes imperative to demonstrate error correction below the QEC threshold: the viability of quantum computing hinges on finding a path to reliable logical qubits.

In Chapter 8 of this thesis, we built a quantum system which is demonstratively below the threshold for bit-flip error correction, for the first time. We have done this using the surface code QEC scheme, a protocol with the most lenient threshold error rate currently known, which also shares a natural synergy with superconducting qubits. We designed the Xmon transmon superconducting qubit, a device that balances the necessary coherence, connectivity, and control needed for a surface code quantum computer. We demonstrated the required physical operations for QEC with the reliability simulated to be at the threshold for the surface

code. Finally, we have used the repetition code, a primitive of the surface code, to suppress environmentally-induced bit-fip errors in a quantum system. This achievement adds significant momentum and credence to qubit technology bringing about the next computing revolution.

Yet, the future holds many challenges before the advent of practical universal quantum computers. Our system contains nine coherent qubits with universal control in a one-dimensional array; estimates for practical quantum computing require millions in a monolithic, two-dimensional array. Scaling current technology to the required number of qubits and more stringent reliability rates will undoubtedly require years of dedicated work and many new inventions. So, let the journey begin.

Appendix A

Supplementary material for Chapter 3: "Coherent Josephson qubit suitable for scalable quantum integrated circuits"

A.1 Device Fabrication

The devices were made in a two-step deposition process. The qubit capacitor, groundplane, readout resonator and control and readout wiring were made in a first, separate deposition step. We used molecular beam epitaxy (MBE) Al deposited on a *c*-plane sapphire substrate. The Al film thickness is approximately 100 nm, deposited at room temperature. The sapphire substrate was cleaned by load-lock outgassing at $200\,^{\circ}\mathrm{C}$, followed by heating to $850\,^{\circ}\mathrm{C}$ in $\sim 10^{-6}$ Torr activated oxygen, identical to the process outlined in Ref. [**?**]. The first Al layer was patterned

by a BCl_3/Cl_2 reactive ion etch.

In the final step, 0.30 x 0.20 μm^2 Al tunnel barriers (30 nm bottom and 100 nm top layer thickness) were made using double-angle shadow evaporation. We used a high vacuum electron-beam evaporator, with a base pressure of approximately 5×10^{-8} Torr. We used the Dolan bridge technique with a poly(methyl methacrylate)/copolymer resist bilayer (approximate thickness: 0.30 and 0.50 μm, respectively), patterned with electron beam lithography. In order to make galvanic contact between the first Al layer and the junction layer, we used a 3 min long Ar ion mill (beam: 400 V, 21 mA; beam width: $\sim$3.2") before shadow-evaporation. Approximately 40 nm was removed from the top resist layer during the ion mill. The junctions were oxidized for 80 mins at 5.0 mBar. Lift-off was done in N-methyl-2-pyrrolidinone at $80\,^{\circ}\mathrm{C}$.

We find that the junctions age very little, the resistance value changes less than 1% over a period of ten days. An optical micrograph of the devices is shown in Fig. A.1.

A.2 Z pulse

In order to quantify the response time of the qubit to a Z pulse, we simultaneously apply a π-pulse with a Gaussian envelope on the XY control and a rectangular wave pattern on the Z control line. The pulse sequence is shown in the left inset in Fig. A.2. We slide the rectangular wave in time ($\Delta\tau$) and offset (ΔZ), while retaining the amplitude of the wave Z_{amp} constant. The excited state population is plotted in Fig. A.2 as a function of time and amplitude.

A cross section of the main figure at $\Delta Z = 0$ (squares) and $\Delta Z = Z_{\mathrm{amp}}$ (circles) is

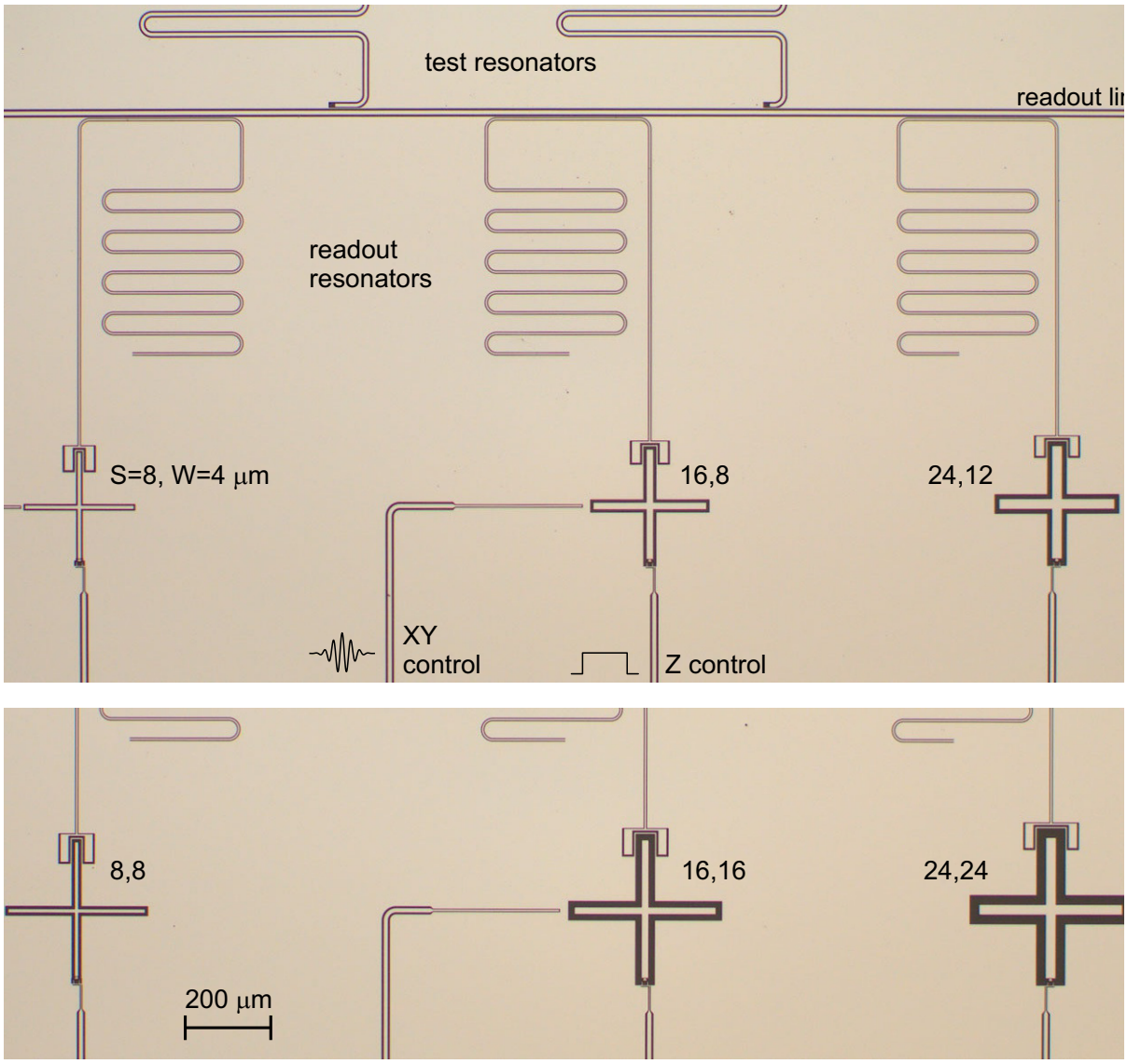


Figure A.1: (Color online) Optical micrograph of the six 'Xmon' qubits on two chips, formed by the Al superconducting (light) and the exposed sapphire substrate (dark). The qubits are capacitively coupled to readout resonators, which couple to a readout line in a frequency multiplexed readout scheme [**?**]. The central linewidth S and gap width W are varied from 8,4 μm to 24,24 μm. The top three Xmons have a single arm length of $L = 130$ μm, the bottom three have $L = 165$ μm. Test resonators provide an independent measurement of the quality factor.

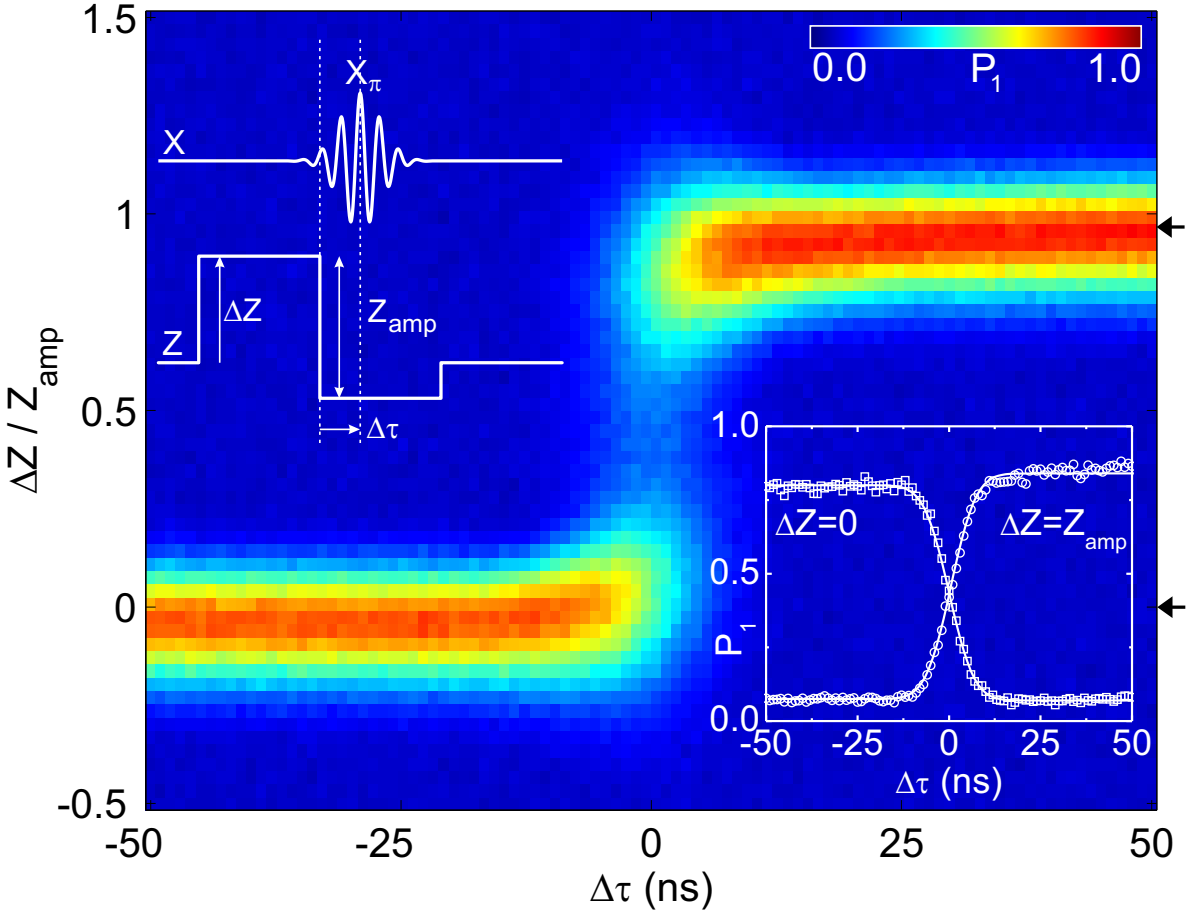


Figure A.2: (Color online) Z pulse shape measurement. The qubit is excited with a π-pulse with 20 ns Gaussian envelope, while a rectangular wave is applied on the Z line. The pulse sequence is shown in the left inset. We use $Z_{\rm amp} = 0.2$ (arb. units) [**?**], which corresponds to a shift in frequency of 73 MHz. The right inset is a cross section at $\Delta Z = 0$ (squares) and $\Delta Z = Z_{\rm amp}$ (circles), as indicated by the arrows (right). The solid lines are fits to partial qubit rotations from a π-pulse with a truncated Gaussian envelope.

shown in the right inset. The measured response can be accurately described by qubit rotation from a partial π-pulse (π-pulse duration: 20 ns), with a truncated Gaussian as envelope (solid lines): with the quantum state given by $\Psi = \cos(\frac{\theta}{2})|0\rangle + \sin(\frac{\theta}{2})|1\rangle$, for the fall at $\Delta Z = 0$: $\theta = \int_{-\infty}^{-\Delta\tau} \sqrt{\pi/2}\exp[-t^2/2\sigma^2]/\sigma dt$. No other time constants are included. We find that the rise at $\Delta Z = Z_{\rm amp}$ is best described when assuming a 0.5 ns delay compared to the fall at $\Delta Z = 0$. We conclude that the qubit frequency is tuned to the desired frequency on a timescale of nanoseconds.

We find the Z control cross-talk between adjacent Xmons to be 1.0-1.5%.

A.3 Qubit decay rate

The frequency-dependent decay rate is displayed in Fig. A.3, replotted from the data in Fig. 4 in the main text.

A.4 Simulation of a qubit-defect system

In order to elucidate coherent as well as incoherent decay of the qubit state, we numerically simulate a qubit-defect system. The system consists of two coupled two-level systems, with coupling strength g and defect energy decay rate $\Gamma_{1,\rm D}$; the qubit is placed on resonance with the defect. The qubit excited state probability is shown in Fig. A.4. For $\Gamma_{1,\rm D} < 4g$, the excitation coherently swaps back and forth between qubit and defect, decaying slowly. When the decay rate exceeds the coupling strength ($\Gamma_{1,\rm D} > 4g$) coherent swapping vanishes and an exponential

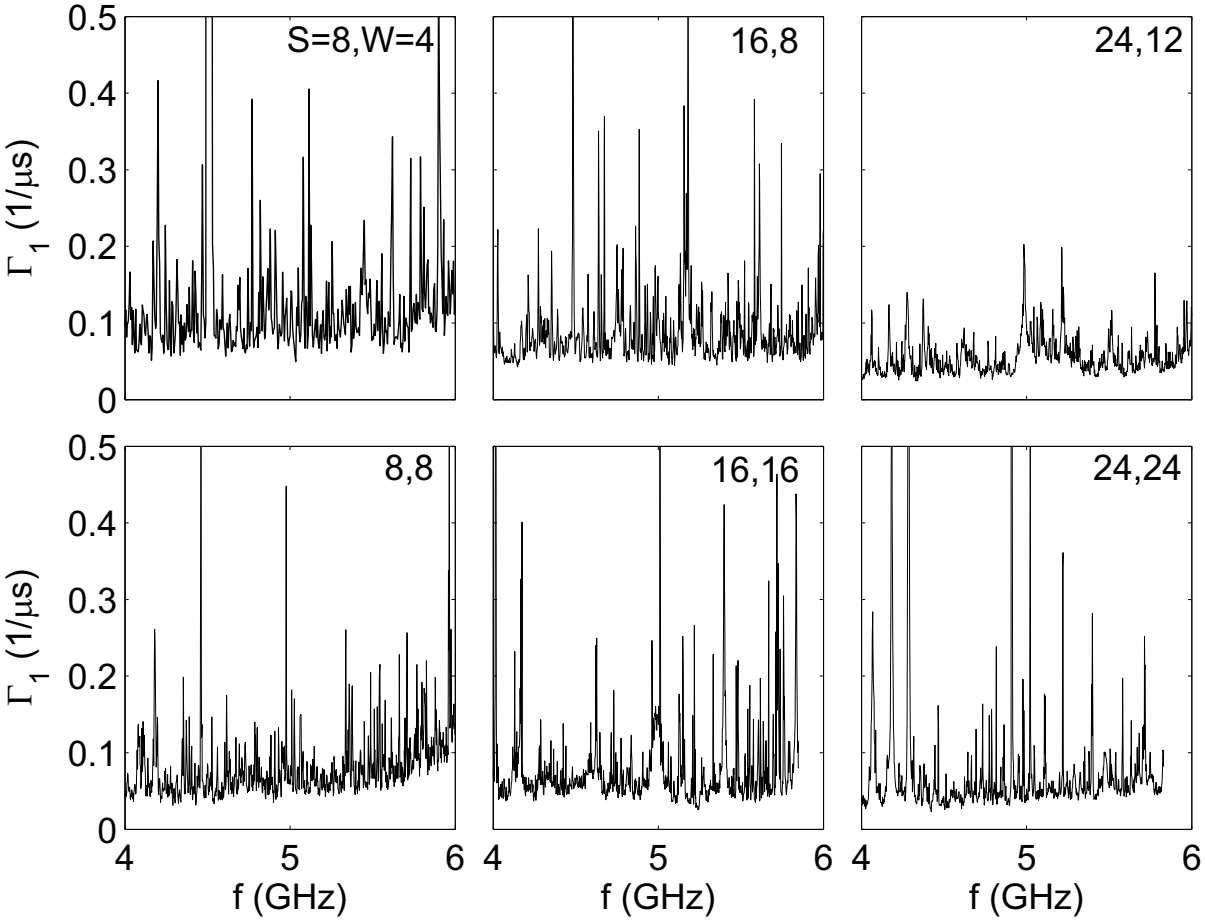


Figure A.3: Frequency dependence of Γ_1 for six qubits with different S and W. The frequency stepsize is 5 MHz for S,W = 8,4 μm and 2 MHz otherwise.

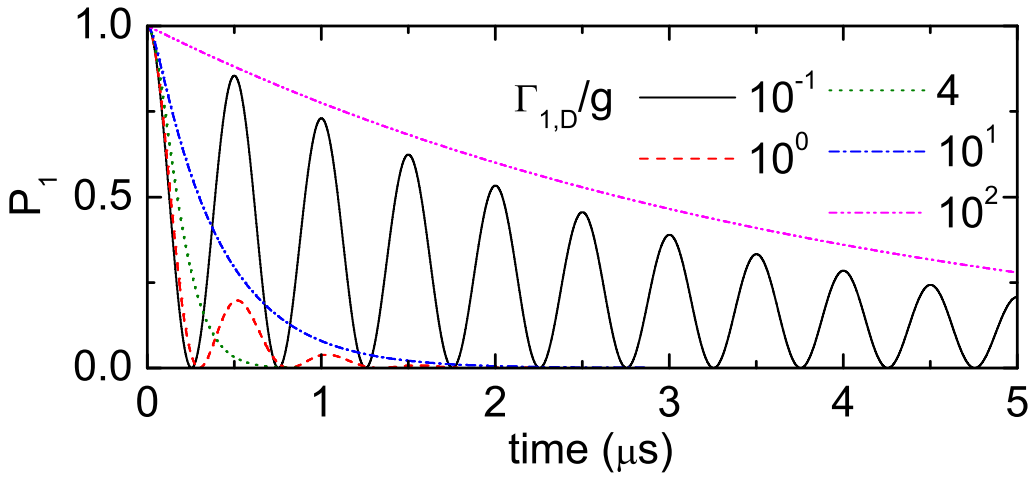


Figure A.4: (Color online) Qubit excited state probability versus time, in a simulation of a qubit-defect system. The qubit is on resonance with the defect, whose energy relaxation time is $\Gamma_{1,\mathrm{D}}$. Here, we used a coupling strength $g/2\pi = 1$ MHz.

decay appears, as the qubit state decays incoherently. The excitation decays most quickly for $\Gamma_{1,\mathrm{D}} = 4g$.

A.5 Analytical expression for loss in a qubit-defect system

Here we derive an analytical expression for the energy loss rate arising from a qubit coupling to a single two-level defect. We consider a system with two coupled two-level systems. We solve the master equation in the Lindblad form,

$$\dot{\rho} = -\frac{i}{\hbar}[H, \rho] + \sum_i \mathcal{D}\left[C_i\right]\rho, \tag{A.1}$$

with $\mathcal{D}[C]\rho = C\rho C^+ - (C^+C\rho + \rho C^+C)/2$ and the Hamiltonian given by

$$H = \hbar\omega_{\mathrm{Q}} a^+a + \hbar\omega_{\mathrm{D}} b^+b + \hbar g(b^+ \otimes a + b \otimes a^+), \tag{A.2}$$

with a and b the lowering operator for qubit and defect, respectively, and ω the angular transition frequency. We model Markovian decoherence through the Lindblad terms: $C_{i=1-4} = \{a\sqrt{\Gamma_{1,\mathrm{Q}}},\ a^+a\sqrt{2\Gamma_{\phi,\mathrm{Q}}},\ b\sqrt{\Gamma_{1,\mathrm{D}}},\ b^+b\sqrt{2\Gamma_{\phi,\mathrm{D}}}\}$, denoting energy and phase relaxation for qubit and defect, respectively. Here we used the number operators a^+a and b^+b to express pure dephasing, and $\Gamma_{1,\mathrm{Q}}$ denotes qubit relaxation. We take $\Gamma_{1,\mathrm{Q}} \ll \Gamma_{1,\mathrm{D}}$. As we are interested in the relaxation of a single excitation, we only consider the states $\{|00\rangle,\ |01\rangle,\ |10\rangle\}$. In the

interaction picture and matrix form the above becomes

$$H = \hbar \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & g \\ 0 & g & \Delta \end{pmatrix}, \tag{A.3}$$

$$\begin{aligned} \dot{\rho} = & -\frac{i}{\hbar}[H, \rho] \\ & -\frac{\Gamma_{1,\mathrm{Q}}}{2} \begin{pmatrix} -2\rho_{22} & \rho_{12} & 0 \\ \rho_{21} & 2\rho_{22} & \rho_{23} \\ 0 & \rho_{32} & 0 \end{pmatrix} - \Gamma_{\phi,\mathrm{Q}} \begin{pmatrix} 0 & \rho_{12} & 0 \\ \rho_{21} & 0 & \rho_{23} \\ 0 & \rho_{32} & 0 \end{pmatrix} \\ & -\frac{\Gamma_{1,\mathrm{D}}}{2} \begin{pmatrix} -2\rho_{33} & 0 & \rho_{13} \\ 0 & 0 & \rho_{23} \\ \rho_{31} & \rho_{32} & 2\rho_{33} \end{pmatrix} - \Gamma_{\phi,\mathrm{D}} \begin{pmatrix} 0 & 0 & \rho_{13} \\ 0 & 0 & \rho_{23} \\ \rho_{31} & \rho_{32} & 0 \end{pmatrix}, \end{aligned} \tag{A.4}$$

with ρ the density matrix, and $\Delta = \omega_{\mathrm{D}} - \omega_{\mathrm{Q}}$.

We are interested in the decay of the qubit excited state probability ρ_{22}, the relevant equations extracted from above are

$$\dot{\rho}_{22} = -ig(\overline{\rho_{23}} - \rho_{23}) - \Gamma_{1,\mathrm{Q}}\rho_{22} \tag{A.5a}$$

$$\dot{\rho}_{23} = -ig(\rho_{33} - \rho_{22}) + i\Delta\rho_{23} - \Gamma\rho_{23} \tag{A.5b}$$

$$\dot{\rho}_{33} = -ig(\rho_{23} - \overline{\rho_{23}}) - \Gamma_{1,\mathrm{D}}\rho_{33} \tag{A.5c}$$

with $\Gamma = \Gamma_{1,\mathrm{D}}/2 + \Gamma_{\phi,\mathrm{D}} + \Gamma_{1,\mathrm{Q}}/2 + \Gamma_{\phi,\mathrm{Q}}$.

In the limit $\Gamma_{1,\mathrm{D}} > g$, $\rho_{33} \approx 0$, and we can approximate the system with two coupled differential equations. Inserting $\rho_{22} = \exp(-\Gamma_1 t)$ and $\rho_{23} = (\beta_r + i\beta_i)\exp(-\Gamma_1 t)$ gives

$$\Gamma_1 = 2g\beta_i + \Gamma_{1,\mathrm{Q}} \tag{A.6a}$$

$$\Gamma_1\beta_r = \Delta\beta_i + \Gamma\beta_r \tag{A.6b}$$

$$\Gamma_1\beta_i = -g - \Delta\beta_r + \Gamma\beta_i, \tag{A.6c}$$

The solution for the qubit energy decay rate Γ_1 in the presence of a two-level defect is (for $\Gamma_{1,\mathrm{D}} > g > \Gamma_{1,\mathrm{Q}}$)

$$\Gamma_1 = \frac{2g^2\Gamma}{\Gamma^2 + \Delta^2} + \Gamma_{1,\mathrm{Q}}. \tag{A.7}$$

We note that the above equation is similar to the low power limit of the two-level system response function, see for example Ref. [**?**].

A.6 Monte Carlo simulation of defects in the Xmon qubit

In order to quantitatively understand the Xmon qubit's energy decay as well as its variation over frequency, we have performed a Monte Carlo simulation for defects in the capacitor. The defect density for AlO_x in tunnel barriers has been established in measurements with phase qubits [**?**], with the distribution over dipole moment p, volume, and frequency given by $\rho_0\sqrt{1 - p^2/p_{\max}^2}/p$, with $\rho_0 \approx 10^2/\mu\mathrm{m}^3/\mathrm{GHz}$, and the maximum dipole moment $p_{\max} =$

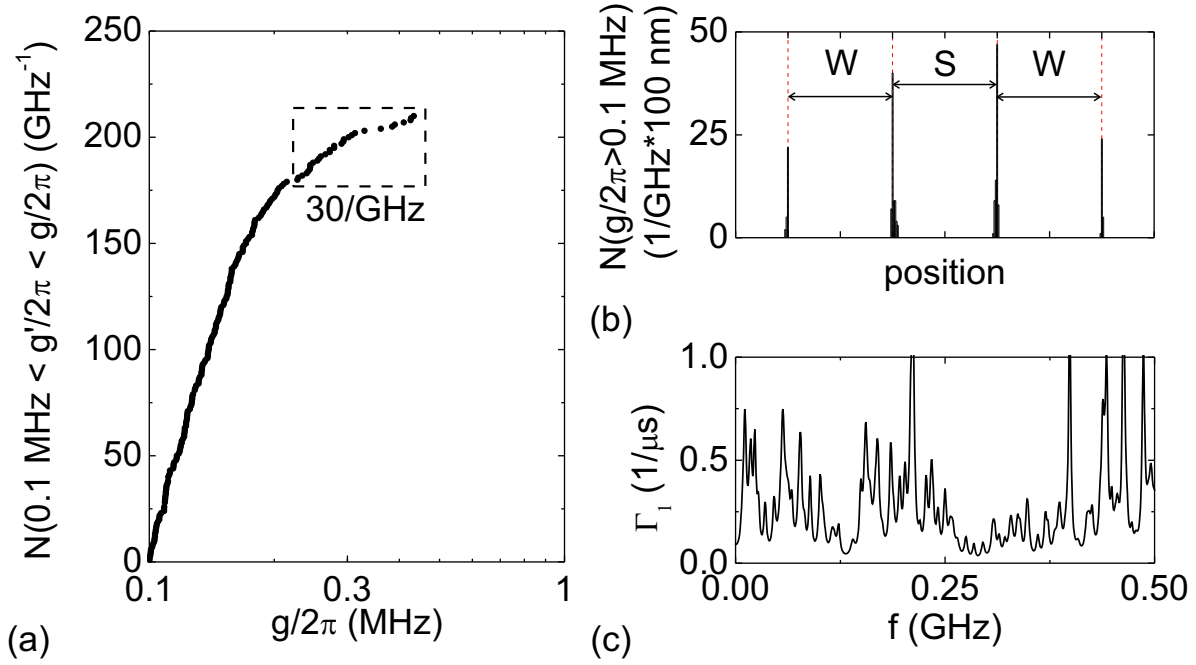


Figure A.5: (Color online) Monte Carlo simulation for defects in the Xmon capacitor, with S=8, W=8 μm and the Xmon arm length L=165 μm. (a) Truncated cumulative distribution of the defect coupling strength. A total of 210 defects/GHz have a coupling strength $g/2\pi > 0.1$ MHz. A total of 30 defects/GHz have a coupling strength $g/2\pi > 0.22$ MHz. (b) Distribution of the defects along position; number of defects in each 100 nm wide section is plotted. The cross section of the capacitor is dashed. The coplanar waveguide geometry as well as the definition of S and W are shown in Fig. 1b in the main text. (c) The simulated decay rate in a 0.5 GHz bandwidth using the distribution in (a,b), and Eq. A.7.

6 D. As the capacitor metal oxide and exposed substrate both consist of Al oxide, we assume that these numbers are a fair representation of the defect density in the qubit capacitor. We only consider defects in the capacitor, the junction is assumed to contain no defects for this simulation.

We randomly place defects in a 3 nm thick dielectric layer ($\epsilon_r = 10$) on the substrate-air and metal-air interfaces (top metal surface as well as the etched edges). The substrate-metal interface is assumed to be thoroughly cleaned [**?**] and to contain no significant defect density. Using a thickness of 2 nm instead of 3 nm for the dielectric layer does not significantly influence the results. Each defect is given a random dipole moment p, such that the distribution over dipole moments matches: $\rho_0\sqrt{1-p^2/p_{\max}^2}/p$, as given by Ref. [**?**]. The coupling strength $g = pE$ is then calculated for each defect using a simulation for the electric fields in our geometry. The results are shown in Fig. A.5.

In Fig. A.5a, we have plotted the truncated cumulative defect distribution over coupling strength. We find that, in a 1 GHz band, 30 defects have a coupling strength of $g/2\pi \gtrsim 0.2$ MHz (square). This simulated value is close to the experimentally observed density of $\sim$30/GHz in Fig. 3 in the main text. These strongly coupled defects are predominantly located within a $\sim$100 nm distance from the etched metal edges, including the exposed substrate surface close to the metal edges and capacitor metal oxide, where the electric fields are largest, see Fig. A.5b.

The simulated qubit decay rate for the same defect distribution is shown in Fig. A.5c in a 0.5 GHz bandwidth. For the defect decay rate we have assumed, for defects with $g/2\pi >$

0.1 MHz, a uniform distribution based on the values extracted from Fig. 3 in the main text: $1/\Gamma_{2,\mathrm{D}} = 50 - 100$ ns ($\Gamma_{2,\mathrm{D}} = \Gamma_{1,\mathrm{D}}/2 + \Gamma_{\phi,\mathrm{D}}$). The simulated decay rate reproduces both the peaks as well as the variation in the background which are observed in the measurement.

Appendix B

Supplementary Information for Chapter 4: "Superconducting quantum circuits at the surface code threshold for fault tolerance"

B.1 Device Fundamentals

B.1.1 Fabrication

The devices are made in a process similar to the fabrication steps outlined in Ref. [**?**], with an important improvement: we have added crossovers to suppress stray microwave chip modes by tying the ground planes together with low impedance connections. Otherwise, the many

control wires in our chip could lead to segmentation of the ground plane, and the appearance of parasitic slotline modes [**?**]. The device is made in a five-step deposition process, (I) Al deposition and control wiring etch, (II) crossover dielectric deposition, (III) crossover Al deposition, (IV) Qubit capacitor and resonator etch, (V) Josephson junction deposition. The qubit capacitor, ground plane, readout resonators, and control wiring are made using molecular beam epitaxy (MBE)-grown Al on sapphire [**?**]. The control wiring is patterned using lithography and etching with a BCl_3/Cl_2 reactive ion etch. A 200 nm thick layer of SiO_2 for the crossover dielectric is deposited in an e-beam evaporator, followed by lift-off. We fabricate crossovers on all the control wiring, using a SiO_2 dielectric that has a non-negligible loss tangent. An in-situ Ar ion mill is used to remove the native AlO_x insulator, after which a 200 nm Al layer for the crossover wiring is deposited in an e-beam evaporator, followed by lift-off. We used 0.9 μm i-line photoresist, lift-off is done in N-methyl-2-pyrrolidone at 80°C. A second BCl_3/Cl_2 etch is used to define the qubit capacitor and resonators; this step is separate from the wiring etch to prevent the sensitive capacitor from seeing extra processing. Lastly, we use electron beam lithography, an in-situ Ar ion mill, and double angle shadow evaporation to deposit the Josephson junctions, in a final lift-off process. See Ref. [**?**] for details.

B.1.2 Coherence Times

Energy relaxation times T_1 of all qubits are shown in Fig. B.1, measured over a frequency range from 4 to 6 GHz. We find typical T_1 values between 20 and 40 μs. Variations in T_1 arise predominantly from the qubit interacting incoherently with weakly coupled two-level defects,

as discussed in Ref. [**?**]. In this previous work we found that larger area (with longer and wider legs) Xmon qubits showed higher T_1 values as well as large, frequency-specific suppressions in the energy coherence: for certain frequencies the T_1 would decrease to values below 10 μs. We attribute these large suppressions to chip modes, arising from imbalances in the microwave control lines, to which the larger Xmon geometries can couple more strongly. The data in Fig. B.1 exhibit fewer of such suppressions; we believe that this improvement is due to the addition of crossovers.

We have investigated the Ramsey dephasing times versus frequency for qubit Q_1. The Ramsey decay envelope is measured by phase tomography (see Ref. [**?**]) and fitted to the function $\exp[-t/T_{\phi,1} - (t/T_{\phi,2})^2]$. Fast dephasing, from white noise as well as energy relaxation, is captured in $T_{\phi,1}$, and slow, Gaussian dephasing is captured in $T_{\phi,2}$. Typical dephasing times are plotted in Fig. B.2. We find a fast dephasing time on the order of 10 μs; this value is below the energy coherence time, and may be due to white noise from the room temperature control electronics. The slow, Gaussian dephasing times are consistent with a $1/f$-spectrum with a spectral density of $S_\Phi(1\text{ Hz}) = 1.1\ \mu\Phi_0/\sqrt{\text{Hz}}$.

Table B.1: Qubit frequencies and nonlinearities ($f_{21} - f_{10}$) at the zero flux bias (degeneracy point) and coupling strengths in MHz. The coupling strength is measured at frequencies between 4.2 and 4.7 GHz. We find a typical next-nearest neighbour coupling of $g/2\pi = 1.3$ MHz, consistent with microwave circuit simulations.

qubits	Q_0	Q_1	Q_2	Q_3	Q_4
f_{10}	5805	5238	5780	5060	5696
nonlinearity	-217	-226	-214	-212	-223
$g_{01}/2\pi$ (4.22 GHz)		27.7			
$g_{12}/2\pi$ (4.70 GHz)			30.8		
$g_{23}/2\pi$ (4.66 GHz)				30.4	
$g_{34}/2\pi$ (4.65 GHz)					30.9

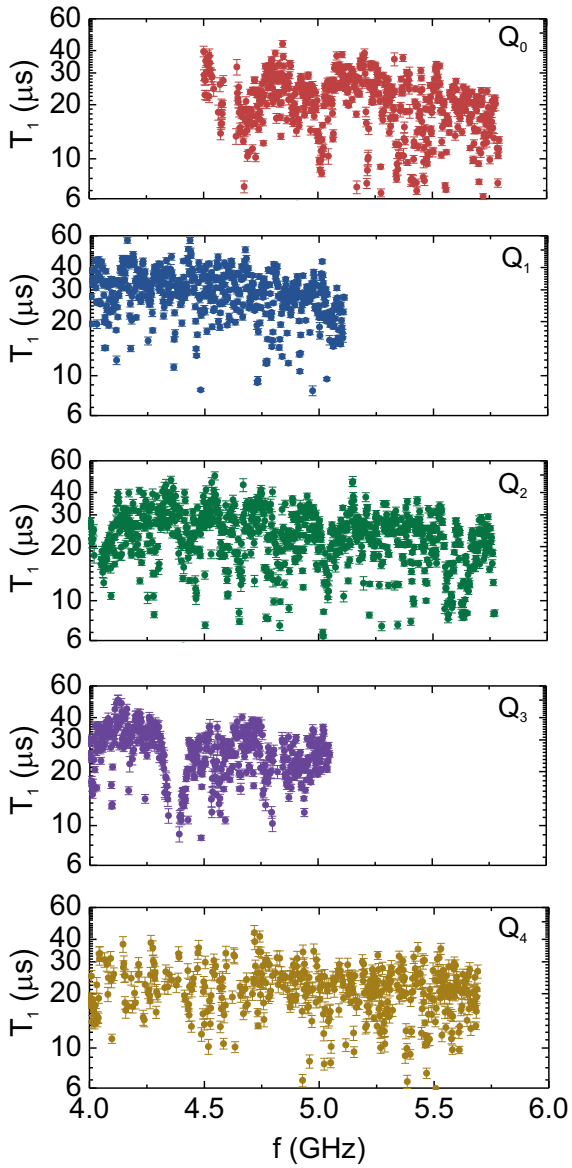


Figure B.1: **Energy Relaxation for Xmon Qubits.** Frequency dependence of T_1 for all qubits. The frequency step size is 2 MHz. The values for T_1 are generally in the 20-40 μs range, we find T_1 values up to 57 μs. The depression at 4.36 GHz in qubit Q_3 is due to a coherently coupled junction defect.

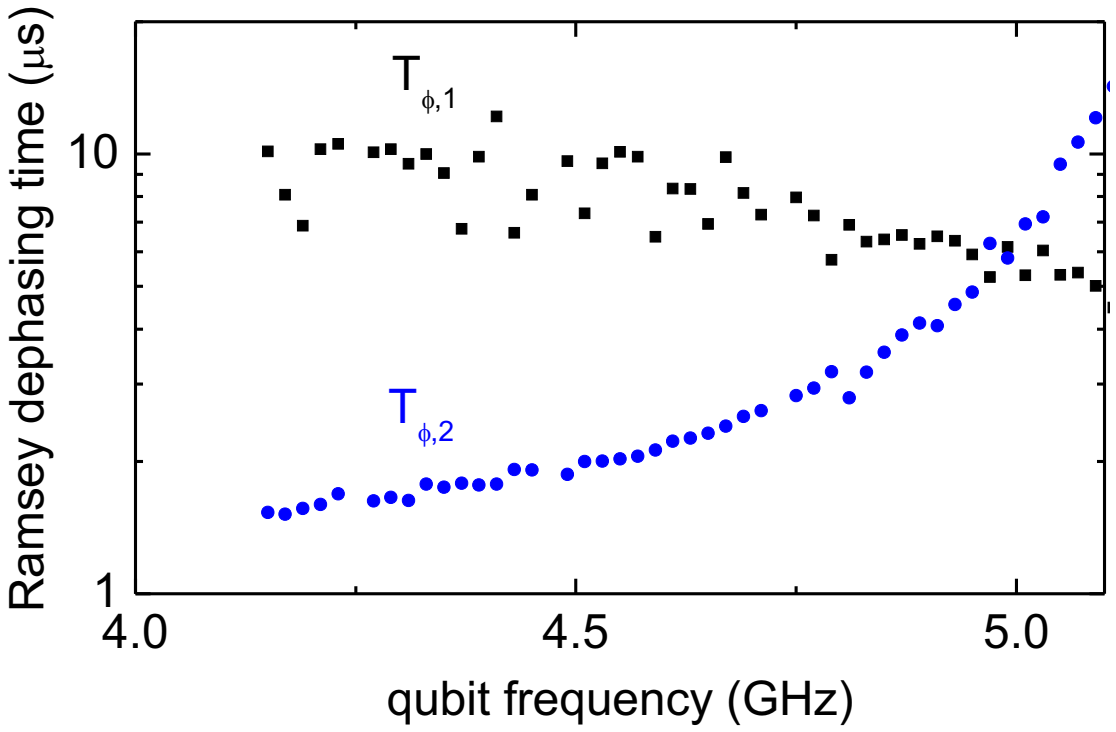


Figure B.2: **Ramsey dephasing.** Frequency dependence of the fast ($T_{\phi,1}$) and slow ($T_{\phi,2}$) Ramsey dephasing times of qubit Q_1. Flux bias points Φ/Φ_0 range from 0.1 to 0.28, and $\delta f_{10}/\delta(\Phi/\Phi_0)$ range from -16 to -50 GHz.

B.1.3 Qubit Frequencies and Coupling

Qubit frequencies and nearest neighbour coupling strengths are listed in Table B.1.

B.1.4 Z Crosstalk

We measure a crosstalk between the frequency Z control lines and qubits that is small, approximately $1 - 2\%$. After adding compensation pulses to orthonormalise the control, we find a remnant crosstalk of below 10^{-4}. The crosstalk matrix M_Φ is shown below, defined as: $\Phi_{\text{actual}} = M_\Phi \Phi_{\text{ideal}}$, with Φ the flux threaded through each qubit's superconducting quantum interference device (SQUID) loop.

$$M_\Phi = \begin{pmatrix} 1.000 & -0.023 & -0.014 & -0.009 & -0.006 \\ 0.019 & 1.000 & -0.022 & -0.011 & -0.007 \\ 0.017 & 0.000 & 1.000 & -0.016 & -0.009 \\ 0.016 & 0.008 & -0.015 & 1.000 & -0.014 \\ 0.013 & 0.014 & -0.016 & -0.010 & 1.000 \end{pmatrix}$$

B.2 Experimental Setup

The wiring diagram and circuit components are shown in Fig. B.3.

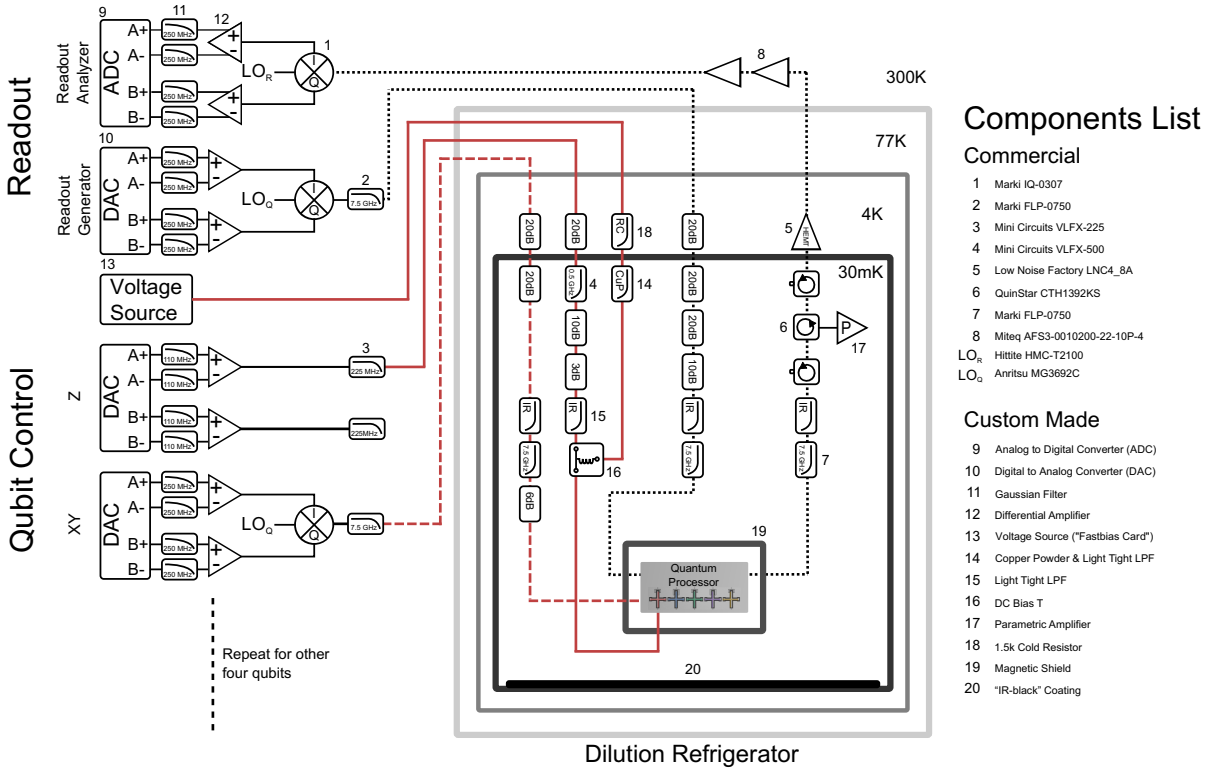


Figure B.3: **Electronics and Control Wiring.** Diagram detailing all of the control electronics, control wiring, and filtering for the experimental setup. Each qubit uses one digital to analog converter (DAC) channel for each of the X, Y, and Z rotations. Additionally, we use a DC bias tee to connect a voltage source to each qubit frequency control line to give a static frequency offset. All five qubits are read out using frequency-domain multiplexing on a single measurement line. The readout DAC generates five measurement tones at the distinct frequencies corresponding to each qubit's readout resonator. The signal is amplified by a wideband parametric amplifier [**?**], a high electron mobility transistor (HEMT), and room temperature amplifiers before demodulation and state discrimination by the analog to digital converter (ADC). All control wires go through various stages of attenuation and filtering to prevent unwanted signals from disturbing the quantum processor. Two local oscillators (LO_Q) are used for qubit XY control, at 4.5 and 5.6 GHz. The readout LO_R is at 6.76 GHz. All LO, DAC, and ADC electronics are locked to a 10 MHz SRS FS725 rubidium frequency standard.

B.3 Flattening the Z response

Imperfections in the frequency control wiring can cause ripples after a pulse. Left unchecked, these can affect gate fidelity significantly, appearing as single qubit phase errors, see Fig. B.4. We employ a two-step procedure to correct for these non-idealities. We first calibrate the room temperature electronics by measuring the unit step (Heaviside step) response at the output of the Z control board.

With the board response corrected by deconvolution, we measure the qubit phase as a function of time $\Delta\tau$ after the end of a unit step. This probes the transfer function of the fridge wiring, contact pads and on-chip control lines. When no unit step is applied, the X/2 pulse rotates the qubit state onto the Y axis. When applying a unit step, deviations in frequency will cause the Bloch sphere vector to deviate from the Y axis. A subsequent Y/2 pulse will make this apparent in the measured excited state probability. We note that this measurement is first order sensitive to small deviations – the difference in probability denotes the phase deviation ($\Delta\phi \approx \Delta P|1\rangle$) – whereas Ramsey and quantum state tomography are second order sensitive: The $\pi/2$ pulses used in tomography project the state onto the Z axis, thus the reconstruction of the phase or state is done from probabilities ($P \approx 1 - \Delta\phi^2/2$) which are second order sensitive to ϕ.

We find that the transfer function can be described by an exponential response with two timescales. Typical values are 100 ns and 5 ns. The longer timescale is consistent with the L/R time arising from the bias tee, with $L \approx 6$ μH and $R = 50$ Ω. We believe that the short timescale arises from reflections. The impulse response of an imperfect wire with reflection r,

placed time T away from the wire's end is $H(\omega) = 1 - r + \sum_{k=1}^{\infty} r^k \exp(-2ik\omega T)$; at low frequencies this can be approximated by the impulse response function $h(t) \propto \exp(-t/2rT)u(t)$. Assuming reflection coefficients on the order of -10 dB and round trip times $2T$ between qubit and mixing plate electronics on the order several ns, the effective decay time $2rT$ is on the order of a few ns.

With the corrections in place, by deconvolving both the board response and fridge wiring, remnant control pulse ripples are suppressed to below 10^{-4}: We find qubit phase deviations consistent with a 30 kHz drift after applying a 0.5 GHz detuning step pulse, see Fig. B.4. The calibrations discussed above are key for obtaining accurate CZ gates [**?**].

B.4 Single qubit and two-qubit gate fidelities of all qubits

A comprehensive listing of all single qubit gate fidelities of all qubits is shown in Table B.2, the gate durations are in Table B.3. A listing of all CZ gate fidelities can be found in Table B.4.

B.5 Scaling up the architecture

In the present experiment, we have demonstrated single- and two-qubit gates at the surface code threshold in a one-dimensional array with five Xmon qubits. Scaling up to a larger one-dimensional array is trivial by adding more qubits on the sides. For scaling up to a two-dimensional grid, we envision an approach with Xmon qubits placed in a square lattice, capacitively coupled to their four nearest neighbours. The Xmon qubit, with the four arms available

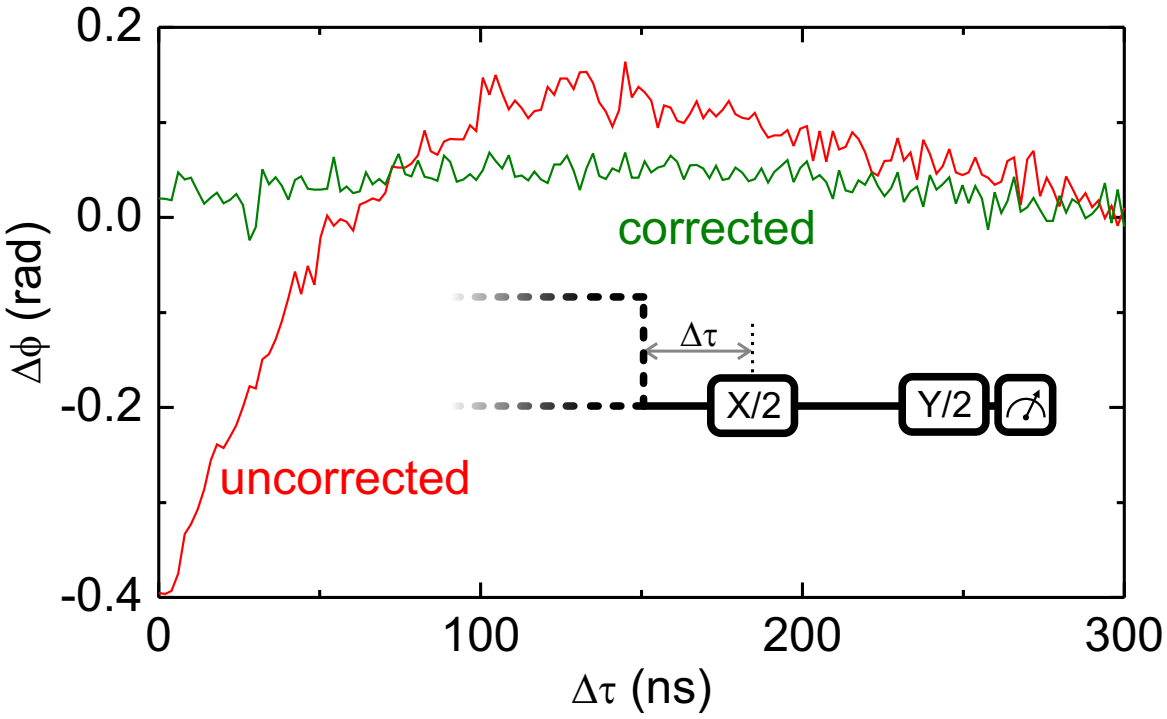


Figure B.4: **Control pulse ripple.** Qubit phase response to a unit step (amplitude: 0.5 GHz) applied to the frequency control line, with and without correction. The pulse sequence is shown in the inset, with the dashed line representing the unit step. With correction, a phase drift of 0.03 rad in 150 ns is observed, consistent with a remnant control pulse ripple of 30 kHz.

Table B.2: Single qubit gate fidelities for all qubits, determined by Clifford-based randomised benchmarking. Averaged over all gates and all qubits we find an average fidelity of 0.9992. The standard deviation is typically $5 \cdot 10^{-5}$. The gate times are between 10 and 20 ns, see Table B.3, except for the composite gates H and 2T, which are twice as long. The idle is as long as the shortest microwave gate (12 ns to 20 ns).

gates	Q_0	Q_1	Q_2	Q_3	Q_4
I	0.9990	0.9996	0.9995	0.9994	0.9991
X	0.9992	0.9996	0.9992	0.9991	0.9991
Y	0.9991	0.9995	0.9993	0.9992	0.9991
X/2	0.9992	0.9993	0.9993	0.9994	0.9993
Y/2	0.9991	0.9993	0.9995	0.9994	0.9994
-X	0.9991	0.9995	0.9992	0.9989	0.9991
-Y	0.9991	0.9995	0.9991	0.9987	0.9991
-X/2	0.9991	0.9992	0.9993	0.9990	0.9995
-Y/2	0.9991	0.9992	0.9995	0.9990	0.9994
H	0.9986	0.9986	0.9991	0.9981	0.9988
Z	0.9995	0.9988	0.9994	0.9991	0.9993
Z/2	0.9998	0.9991	0.9998	0.9995	0.9996
2T [1]		0.9989	0.9994	0.9989	0.9990
average over gates	0.9992	0.9992	0.9994	0.9991	0.9992
average over qubits			0.9992		

Table B.3: Single qubit gate times in ns.

gates	Q_0	Q_1	Q_2	Q_3	Q_4
XY axes π rotations	20	20	12	18	12
XY axes $\pi/2$ rotations	20	20	12	12	12
Z axis π, $\pi/2$, $\pi/4$ rotations	10	10	10	10	10
I	20	20	12	12	12
H	40	40	24	30	24
2T	20	20	20	20	20

Table B.4: CZ gate fidelities for all qubit pairs, determined by Clifford-based randomised benchmarking. Gate times are between 38 and 45 ns; Q_0-Q_1: 45 ns, Q_1-Q_2: 43 ns, Q_2-Q_3: 43 ns, Q_3-Q_4: 38 ns.

qubits	Q_0	Q_1	Q_2	Q_3	Q_4
$CZ_{Q_0-Q_1}$	0.9924 ± 0.0005				
$CZ_{Q_1-Q_2}$		0.9936 ± 0.0004			
$CZ_{Q_2-Q_3}$			0.9944 ± 0.0005		
$CZ_{Q_3-Q_4}$				0.9900 ± 0.0006	

for coupling, has been designed with this task in mind. The qubits are placed in one layer, and in the envisioned approach control and readout are placed on other planes. Methods to connect these planes are known: Bump bonds and vias allow for making galvanic connections between multiple layers, and capacitive coupling between planes is made easy by the presence of a dielectric substrate. While this approach is conceptually straightforward, the required microfabrication technology will need to be developed further. A schematic of the proposed approach is shown in Fig. B.5.

While it is straightforward to lithographically define more qubits on a chip, the crucial challenge in scaling up to a two-dimensional array – with added wiring and readout – is to maintain the high coherence and gate fidelity, without sacrificing addressability of individual qubits. We believe that this will guide the approach to and the development of the required microfabrication. We are optimistic about scaling up, as coherence was maintained in moving from isolated qubits (in Ref. [**?**]) to the one-dimensional array here.

B.6 Verifying experimental fidelities are at the surface code threshold

We shall simulate inceasing size versions of Fig. B.5. Note that the single-qubit and two-qubit gate physics is identical in this architecture, justifying the use of the reported experimental fidelities. One additional feature in 2-D is the need to use a periodic array of frequencies, leading to interactions between qubits with the same frequencies and driving of the entire set

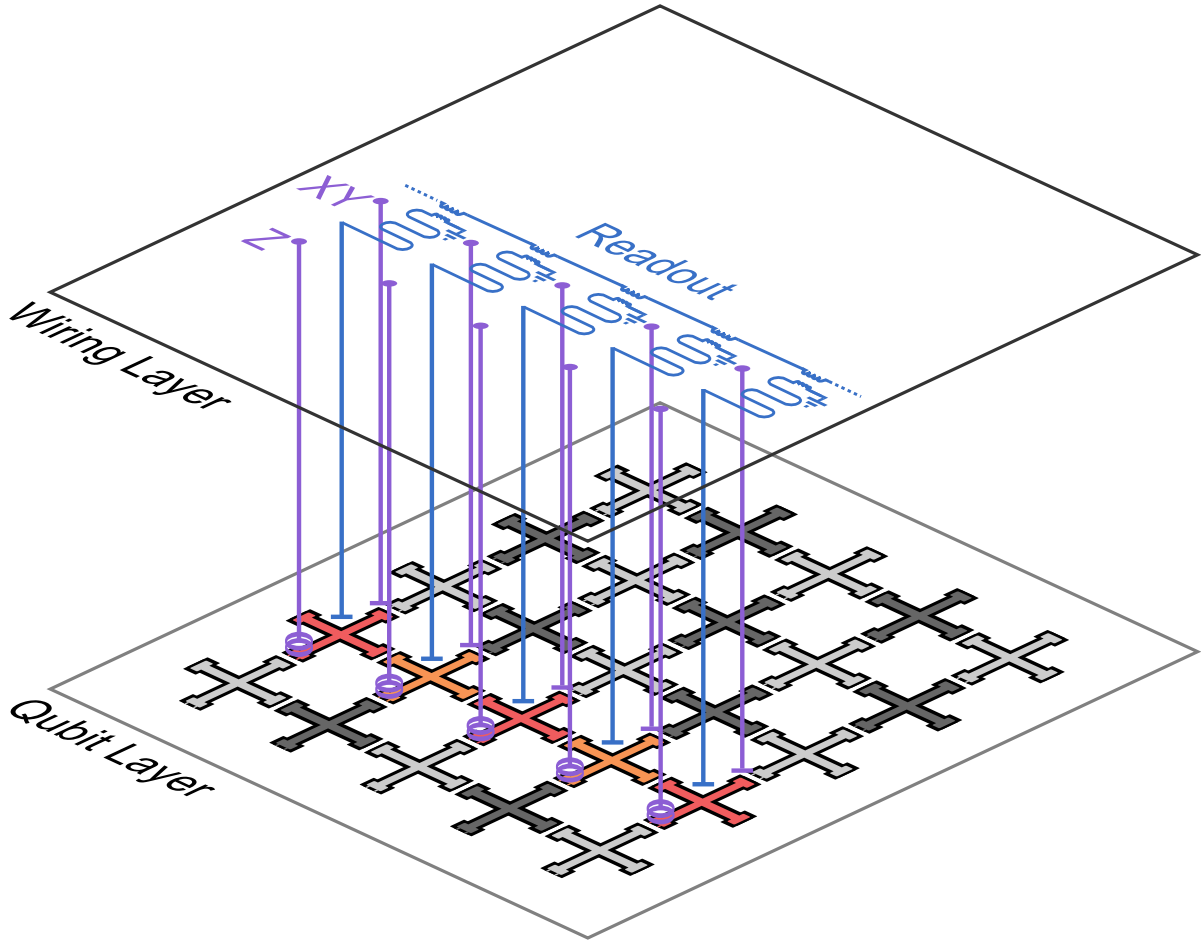


Figure B.5: **Schematic of the proposed approach to scaling up.** The Xmon qubits are placed in a checkerboard pattern in a single layer, with data qubits (light) and measurement qubits (dark). Control, readout, and wiring are placed in other layers; connections between the layers are made by capacitive coupling, and by bump bonds and vias for galvanic connections. The row with the five Xmon qubits coloured red and orange is the experiment we have implemented (see Fig. 4.1), and can clearly be seen as a building block for the two-dimensional array.

of qubits with a given frequency when any given qubit is manipulated. The magnitude of the interactions and driving crosstalk will only decay quadratically with qubit separation.

We have simulated these crosstalk effects in detail and have found them to be surprisingly tolerable, as such errors are only two-body correlated at worst, and non-neighbouring, leading to information local to both halves of the error, and efficient correction [**?**]. We shall therefore neglect crosstalk in the simulations of the section to focus the analysis on the fidelity of the reported gates.

Nominally, the threshold fidelity of the surface code is 0.99 [**?**], provided one assumes there is no leakage, the two-qubit interaction is the dominant source of error, and gates can be performed perfectly in parallel. The physical device described in this work has complex behavior outside these assumptions, necessitating a device-specific calculation of the surface code threshold fidelity.

When a CZ gate is applied, no qubit neighbouring either of the qubits involved in the CZ can be involved in their own CZ gate. We have devised a 16 step CZ application pattern that accounts for these parallelism constraints and still measures all stabilizers. The longest measured CZ time of 45 ns will be used. Furthermore, the CZ gate, which is always applied between one measurement qubit and one data qubit, has a small amount of leakage ($< 0.2\%$) on the measurement qubit, but practically negligible leakage on the data qubit. We shall neglect this small amount of measurement qubit leakage. Methods of coping with leakage in topological codes are known [**?**].

Measurement with fidelity 0.99 in 200 ns and initialization with fidelity 0.99 in 50 ns will

Table B.5: CZ gate error budget, including the contribution to the total error in percent.

Decoherence (55%)	Q_2	≥ 0.0017 (24%)
	Q_3	0.0022 (31%)
Control (45%)	single qubit phase error	≤ 0.0017 (24%)
	state leakage	0.0015 (21%)

be assumed [**?**]. Y/2 gates will be used instead of Hadamard gates, with the slowest 20 ns time assumed and an average fidelity of 0.9992 (calculated only from the slower Y/2 gates) assumed. An identity error of 0.05% per 10 ns will be assumed, consistent with experimental data.

Detailed simulations of 5x5, 9x9, and 13x13 qubit arrays with the above parameters have been performed making use of the latest correction techniques [**?**]. The logical error rate was found to be the same in all cases, justifying our claim of a device with parameters at the surface code threshold.

B.7 CZ gate error budget

We experimentally measure the three predominant error mechanisms of the CZ gate: 2-state leakage, decoherence and control error. 2-state leakage is measured using the same technique as outlined in [**?**]. The system is initialised in the $|11\rangle$-state followed by two CZ gates. As the time between these two gates is varied, we measure an interference pattern in the probability where the magnitude is proportional to the $|02\rangle$ state leakage, see Fig. B.6a. The error is given by $\Delta P/4$, with ΔP the peak-to-peak difference in probability [**?**]. We also see additional interference patterns that come from imperfect $|11\rangle$ preparation at the beginning of

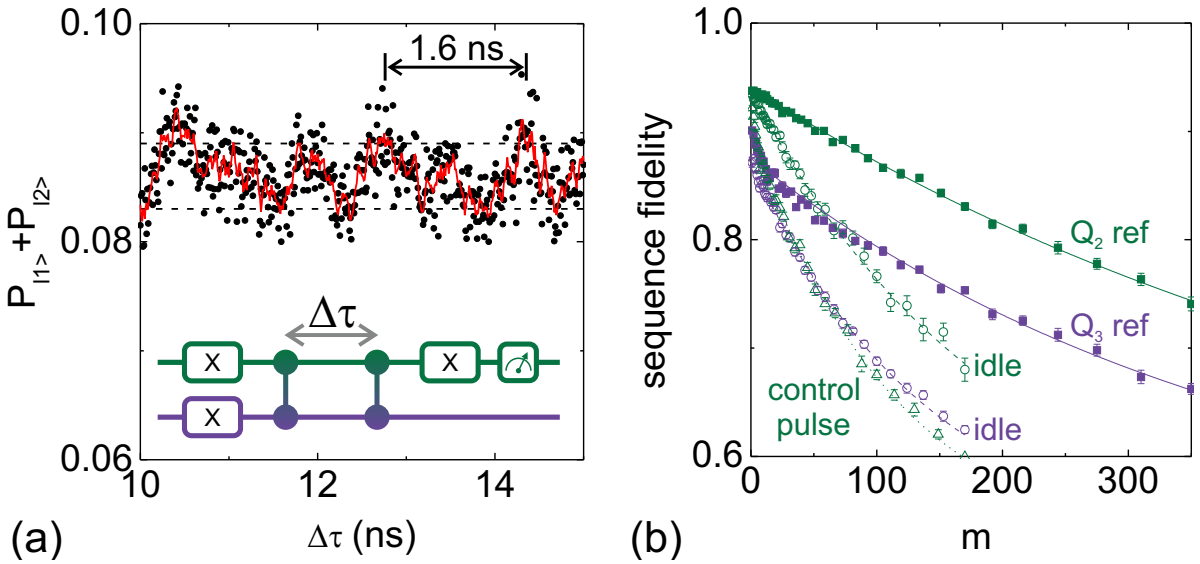


Figure B.6: **CZ error budget.** (a) In the Ramsey error filter technique an interference pattern arises in the measured probability (black dots) whose magnitude is proportional to the $|02\rangle$-state leakage. The data are smoothed (red) for enhanced visibility. The frequency of these oscillations (indicated by the arrow, 1.6 ns) is the idling frequency difference between qubits (800 MHz), minus the nonlinearity (200 MHz) as we are measuring the crossing between $|11\rangle$ and $|02\rangle$. Other frequencies are believed to arise from improper $|11\rangle$ state preparation. (b) Randomised benchmarking sequence fidelity for qubits Q_2 and Q_3. Decoherence is quantified by idling for the same duration as the CZ gate. Controls error can be identified by applying the control pulse on Q_2, without doing a full CZ by detuning Q_3. For the randomised benchmarking data of the CZ (not shown): $r_{\mathrm{ref}} = 0.0198$ and $r_{\mathrm{CZ}} = 0.0269$.

the sequence. We note that leakage occurs predominantly in the qubit which undergoes the frequency trajectory.

We measure the decoherence contribution from each qubit by performing interleaved randomised benchmarking with an idle of the same duration as the CZ gate, see Fig. B.6b. The contribution to error from the waveform is measured by interleaved randomised benchmarking on the waveform for the CZ gate alone, with a slightly lower amplitude to avoid interactions with the other qubit. We treat this as a single-qubit phase gate. With the idle error measured, we can separate out decoherence and single qubit phase error. Because we are detuning the qubit down in frequency to a part of the spectrum where it is more sensitive to flux noise, inducing more dephasing, the single qubit phase error is an upper bound. With these experiments we can construct an error budget for all of the dominant error mechanisms, as seen in Table B.5.

B.8 Quantifying XY control crosstalk using simultaneous randomised benchmarking

Addressability, the ability to individually control a single qubit without affecting neighbouring qubits, is of great importance when building a multi-qubit system. In our five Xmon qubit processor the addressability is mostly compromised in three ways: Z control crosstalk, microwave XY control crosstalk, and off-resonant qubit-qubit coupling. Z crosstalk can be reduced to below the 10^{-4} level. Microwave XY crosstalk becomes a problem if a qubit's control pulses perform rotations on a neighbouring qubit. Off-resonant qubit-qubit coupling will very slowly

perform a CZ gate between the qubits, potentially causing unwanted phase shifts with rate Ω_{ZZ},

$$\Omega_{ZZ} = -\frac{2g^2(\eta_1 + \eta_2)}{(\Delta - \eta_1)(\Delta + \eta_2)}, \tag{B.1}$$

with η_1 and η_2 the qubit nonlinearities, and Δ the difference in qubit frequencies.

We performed crosstalk characterisation on nearest neighbour and next-nearest neighbour qubits. Nearest neighbours are far detuned (> 800 MHz), hence the microwave XY crosstalk is expected to be negligible, but the off-resonant CZ interaction may be non-negligible. Next-nearest neighbors have a much smaller coupling ($g/2\pi = 1.3$ MHz), but are only detuned by 100-400 MHz; hence both the off-resonant CZ as well as microwave XY crosstalk may be detrimental. We investigate these mechanisms by using the simultaneous randomised benchmarking techniques outlined in Ref. [**?**]. We can single out errors that come from poor addressability by performing randomised benchmarking on each qubit individually, and operating both qubits simultaneously.

The randomised benchmarking data are shown in Fig. B.7. We can determine the effect of controlling qubit Q_3 on Q_2, by first benchmarking qubit Q_2 individually ($I \otimes C_1$, green open squares), and benchmarking both qubit Q_2 and Q_3 simultaneously, and tracing out the contribution of Q_3 ($C_1 \otimes C_1$, green full squares). The decay for both traces is virtually indistinguishable, the added error is below 10^{-4}. Likewise, we find that the effect on Q_3 of controlling Q_2 simultaneously leads to an added error per Clifford of $2 \cdot 10^{-4}$. For next nearest neighbours, we find added errors per Clifford of $1 \cdot 10^{-4}$ and $2 \cdot 10^{-4}$. For both the nearest neighbour

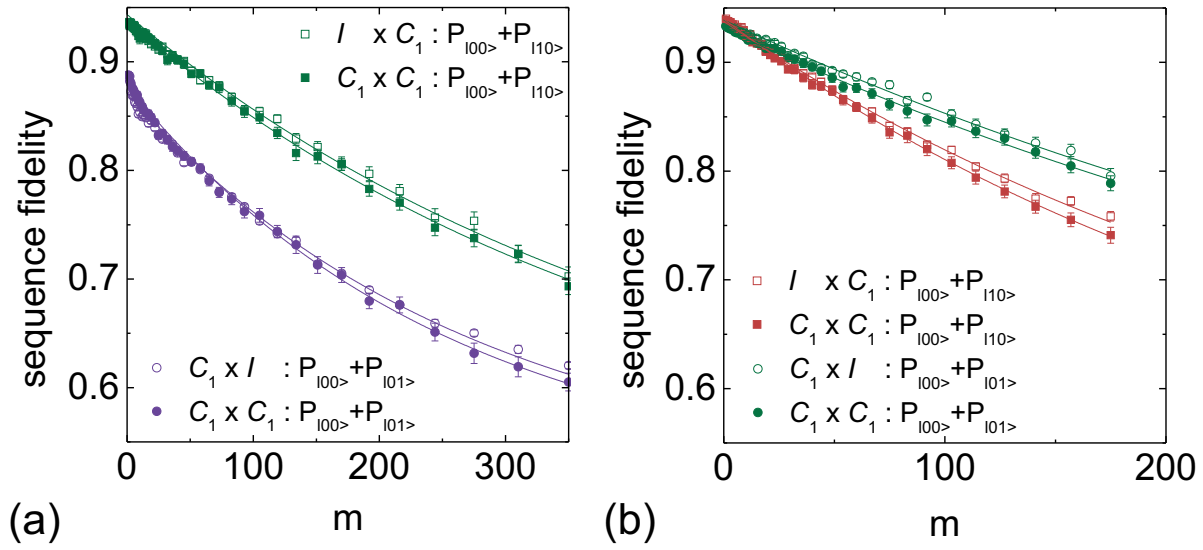


Figure B.7: **Simultaneous randomised benchmarking of nearest and next nearest neighbours.** (a) Benchmarking the effect of Q_2 on Q_3 and vice-versa ($f_{Q_2} = 5.72$ GHz, $f_{Q_3} = 4.67$ GHz). The sequence fidelities are shown for operating Q_2 individually ($I \otimes C_1$, green open squares), Q_3 individually ($C_1 \otimes I$, purple open circles), and Q_2 and Q_3 simultaneously ($C_1 \otimes C_1$, full symbols). By tracing out one qubit, its effect on the other qubit becomes apparent: the errors per Clifford are: r_{Q2}=0.0011, $r_{Q2|Q3}$=0.0012, r_{Q3}=0.0018, $r_{Q3|Q2}$=0.0020. (b) Benchmarking of Q_0 and Q_2 ($f_{Q_0} = 5.30$ GHz, $f_{Q_2} = 5.72$ GHz). The errors per Clifford are: r_{Q0}=0.0016, $r_{Q0|Q2}$=0.0018, r_{Q2}=0.0011, $r_{Q2|Q0}$=0.0011. Note that the errors per Clifford are consistent with the average gate fidelities in Table B.2: for Q_2, the average gate fidelity is $1 - r_{Q2}/1.875$=0.9994. Coupling strengths can be found in Table B.1.

and next-nearest neighbours the added error per Clifford of operating them simultaneously is $< 2 \cdot 10^{-4}$, the inferred added error per single qubit gate is therefore $< 10^{-4}$. We conclude that XY control crosstalk is a minor error mechanism, enabling a high degree of addressability in this architecture.

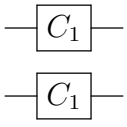
B.9 Generation of the Clifford groups

B.9.1 Single qubit Clifford group C_1

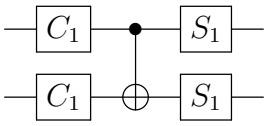
The single qubit Clifford group C_1 is the group of 24 rotations which preserve the octahedron in the Bloch sphere. We implement the group using microwave pulses only, decomposed into rotations around the X and Y axes using the generators: {I, ±X/2, ±Y/2, ±X, ±Y}, as summarised in Table B.6. The average number of single qubit gates per single qubit Clifford is 1.875.

B.9.2 Two qubit Clifford group C_2

Using the single qubit Cliffords, we can construct the two qubit Clifford group C_2 following Ref. [**?**]. This group has four classes: the single qubit class, the CNOT-like class, the iSWAP-like class, and the SWAP-like class. The CNOT and SWAP-like class are terminated with a gate from the 3-element group S_1, as described in Table B.7. The single qubit class has $24^2 = 576$ elements:



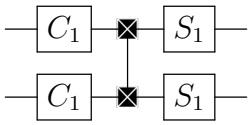
The CNOT-like class has $24^2 \times 3^2 = 5184$ elements.



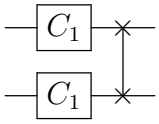
The iSWAP-like class also has 5184 elements,

Table B.6: The 24 single qubit Cliffords written in terms of the physical microwave gates applied in time. The Paulis and $2\pi/3$ rotations form the tetrahedron symmetry group. Here, X/2 denotes a $\pi/2$ rotation over the X axis, with unitary $R_{\mathrm{X}}(\pi/2) = \exp(-i\pi\sigma_{\mathrm{X}}/4)$.

Single qubit Cliffords			
Paulis	I		
	X		
	Y		
	Y,	X	
$2\pi/3$ rotations	X/2,	Y/2	
	X/2,	-Y/2	
	-X/2,	Y/2	
	-X/2,	-Y/2	
	Y/2,	X/2	
	Y/2,	-X/2	
	-Y/2,	X/2	
	-Y/2,	-X/2	
$\pi/2$ rotations	X/2		
	-X/2		
	Y/2		
	-Y/2		
	-X/2,	Y/2,	X/2
	-X/2,	-Y/2,	X/2
Hadamard-like	X,	Y/2	
	X,	-Y/2	
	Y,	X/2	
	Y,	-X/2	
	X/2,	Y/2,	X/2
	-X/2,	Y/2,	-X/2



Finally the SWAP-like class, with 576 elements, is given by



bringing the full size of the two-qubit Clifford group to 11520. An approach using a reduced set of gates was used in Ref. [**?**].

Here, we rewrite the two-qubit Cliffords in terms of the CZ entangling gate. We rewrite the CNOT, iSWAP and SWAP in terms of the CZ:

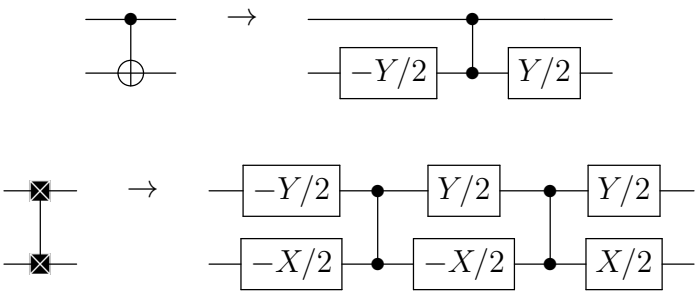
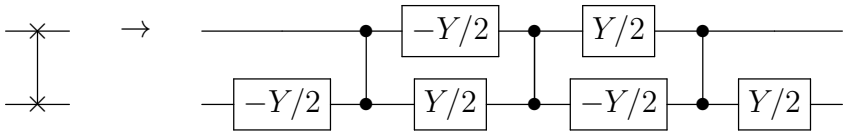
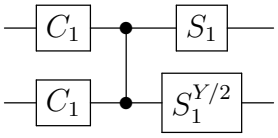


Table B.7: The S_1 sets written in terms of physical gates in time; these are elements of the single qubit Clifford group, and therefore physically implemented in the same way.

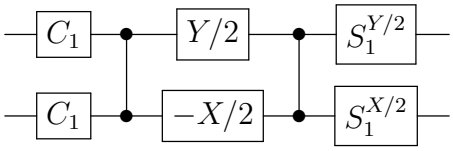
S_1	I Y/2, -X/2,	X/2 -Y/2	
$S_1^{X/2}$	X/2 X/2, -Y/2	Y/2,	X/2
$S_1^{Y/2}$	Y/2 Y, -X/2,	X/2 -Y/2,	X/2



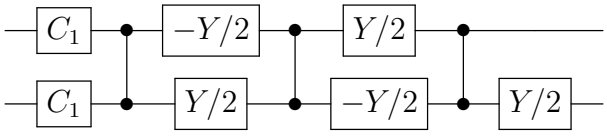
As the single qubit gates preceeding the entangling operation (CZ gate) can be absorbed into C_1, and the final single qubit gates can be absorbed into S_1 (see Table B.7), we have for the CNOT-like class,



the iSWAP-like class,



and the SWAP-like class,



The average number of gates for C_2 is 1.5 CZ gates and 8.25 single qubit gates. For the idle, we wait as long as the shortest single qubit gate. For a single qubit gate time of 20 ns and a CZ gate time of 40 ns, the average duration of C_1 is 37.5 ns, and C_2 is 160 ns.

The Clifford group is a 2-design. A set of unitaries $\{U_k\}_{k=1}^K$ is a 2-design if and only if [**?**]

$$\sum_{k,k'=1}^{K} |\mathrm{Tr}\left(U_{k'}^{\dagger} U_k\right)|^4 / K^2 = 2. \tag{B.2}$$

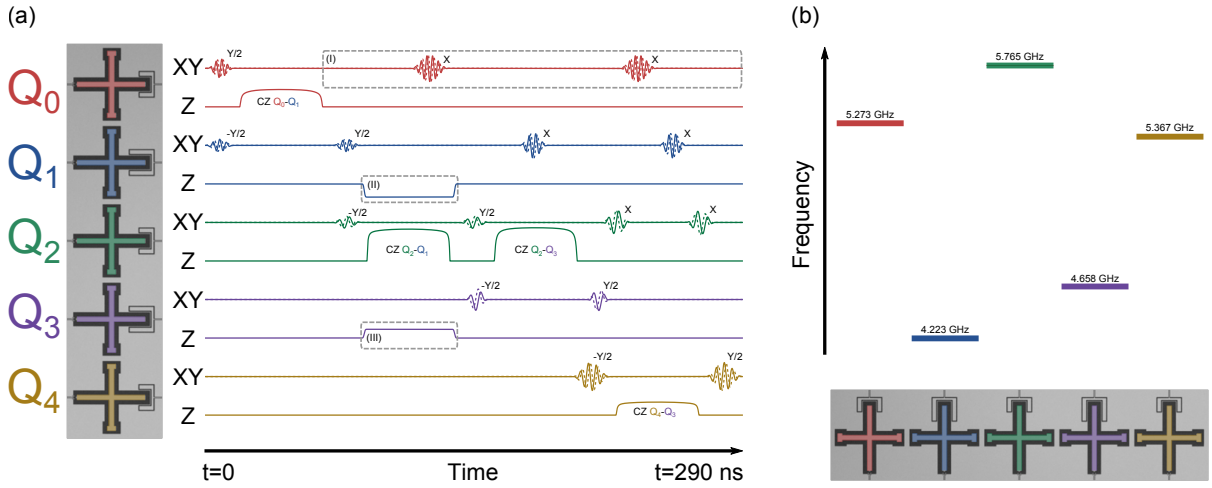


Figure B.8: **Pulse sequence for generating the $N = 5$ GHZ state and frequency diagram.** (a) The control signals for all five qubits. The algorithm consists of a Y/2 pulse on Q_0 followed by successive CNOT gates (implemented here with a CZ gate and -Y/2, Y/2 gates applied to the target) on each progressive pair of qubits in the array. The highlighted region (I) shows Hahn spin echo pulses X applied to Q_0 to suppress dephasing while idling. Spin echo pulses are also applied to Q_1 and Q_2. (II) We detune Q_1 to bring it closer in frequency to Q_2 for the CZ gate. (III) Simultaneously with the Q_1-Q_2 entangling operation, we detune Q_3 away to allow for selective entanglement. (b) The frequency diagram shows the idling frequencies for all qubits, and is one of the operating modes of the quantum processor.

As a consistency check, we have verified that the single and two-qubit Cliffords we generate are indeed a 2-design with the above equation.

B.10 Estimating the Error Per Clifford

Here, we connect the error per Clifford r to the errors of the single and two-qubit gates, measured when performing randomised benchmarking. This shows the physical significance of the error per Clifford, and is an important self-consistency check. We can give an estimate for the error per Clifford by determining the average number of single and two-qubit gates that go into a Clifford, combined with the single and two-qubit gate fidelities that we measure using

interleaved randomised benchmarking. We assume that gate errors are small and uncorrelated, such that adding error when composing gates is a good approximation.

B.10.1 Single qubit Clifford group C_1

There are 45 single qubit gates used across 24 Cliffords. With the assumption that all single gates have the same error, the average error per Clifford is

$$r_{C_1} = 1.875 r_{\mathrm{SQ}}, \tag{B.3}$$

with r_{SQ} the average single-qubit gate error.

B.10.2 Two qubit Clifford group C_2

The four classes of two-qubit Cliffords are composed from the two-qubit CZ gate, and the single-qubit gate sets C_1, S_1, $S_1^{Y/2}$, and $S_1^{X/2}$. The respective errors are given by: $r_{S_1} = 5r_{\mathrm{SQ}}/3$, $r_{S_1^{X/2}} = 5r_{\mathrm{SQ}}/3$, $r_{S_1^{Y/2}} = 2r_{\mathrm{SQ}}$.

We now derive the average gate composition for the two-qubit Cliffords. For the single-qubit class:

$$r_{C_1 \otimes C_1} = \frac{90}{24} r_{\mathrm{SQ}}. \tag{B.4}$$

CNOT-like class:

$$r_{\text{CNOT}} = r_{\text{CZ}} + \frac{89}{12} r_{\text{SQ}}. \tag{B.5}$$

iSWAP-like class:

$$r_{\text{iSWAP}} = 2r_{\text{CZ}} + \frac{113}{12} r_{\text{SQ}}. \tag{B.6}$$

SWAP-like class:

$$r_{\text{SWAP}} = 3r_{\text{CZ}} + \frac{35}{4} r_{\text{SQ}}. \tag{B.7}$$

The error per Clifford for C_2 is then given by

$$r_{C_2} = \frac{576}{11520} r_{C_1 \otimes C_1} + \frac{5184}{11520} r_{\text{CNOT}} + \tag{B.8}$$

$$\frac{5184}{11520} r_{\text{iSWAP}} + \frac{576}{11520} r_{\text{SWAP}} \tag{B.9}$$

$$= \frac{3}{2} r_{\text{CZ}} + \frac{33}{4} r_{\text{SQ}}. \tag{B.10}$$

And the error per two-qubit Cliffords interleaved with a CZ is

$$r_{C_2+\text{CZ}} = \frac{5}{2} r_{\text{CZ}} + \frac{33}{4} r_{\text{SQ}}. \tag{B.11}$$

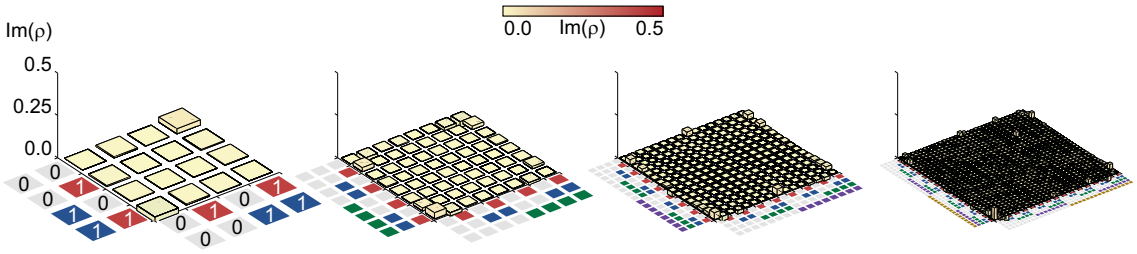


Figure B.9: **Quantum state tomography of the GHZ states: Imaginary parts.** Imaginary parts of the density matrices ρ for the $N = 2$ Bell state and the $N = 3, 4$ and 5 GHZ states. For clarity, the same scale as for the main Letter is used. $|\mathrm{Im}\rho|$ is below 0.03, 0.04, 0.04 and 0.07 for $N = 2$ to 5, respectively.

B.10.3 Comparison to Experiment

Using these simple formulas, we find that our randomised benchmarking data are self-consistent. Using reasonable values of 0.001 and 0.006 for the single and two-qubit gates respectively, we calculate $r_{C_2} = 0.0173$, which is close to the experimental value of $r_{\mathrm{ref}} = 0.0189$ in Fig. 3 in the main Letter; for the interleaved case the calculated value of $r_{C_2+\mathrm{CZ}} = 0.0233$ is close to the experimental value of 0.0244 as well.

B.11 $N = 5$ GHZ state Pulse sequence

The pulse sequence for the algorithm to construct the five qubit GHZ state is shown in Fig. B.8a. We use Hahn spin echoes on idling elements to suppress slow dephasing ($T_{\phi,2}$). The frequency diagram for the qubits is shown in Fig. B.8b. Nearest neighbour qubits are detuned by 0.7 to 1.5 GHz, next-nearest neighbours are detuned by 0.4 to 0.5 GHz.

B.12 Quantum state tomography

The density matrices of the $N = 2$ Bell and $N = 3, 4, 5$ GHZ states are characterised using quantum state tomography. After state preparation, gates from $\{$ I, X/2, Y/2, X $\}^{\otimes N}$ are applied; with the measured probabilities the state can then be reconstructed. We use quadratic maximum likelihood estimation, using the MATLAB packages SeDuMi and YALMIP, to extract the density matrix while constraining it to be Hermitian, unit trace, and positive semidefinite; the estimation is overconstrained. Non-idealities in measurement and state preparation are suppressed by performing tomography on a zero-time idle [**? ?**]. We note that tomography is only "as good as" the tomography pulses, which have an average fidelity above 0.999. Fidelities and uncertainties correspond to the mean and standard deviation of 10 measurements, consisting of 10^4 ($N = 2, 3$) or $6 \cdot 10^3$ ($N = 4, 5$) repetitions each. The density matrices plotted in the main Letter are constructed by averaging all measured probabilites, effectively using 10^5 ($N = 2, 3$) or $6 \cdot 10^4$ ($N = 4, 5$) repetitions.

The imaginary part of the density matrices (ρ) is plotted in Fig. B.9. The Pauli operator representation is shown in Fig. B.10.

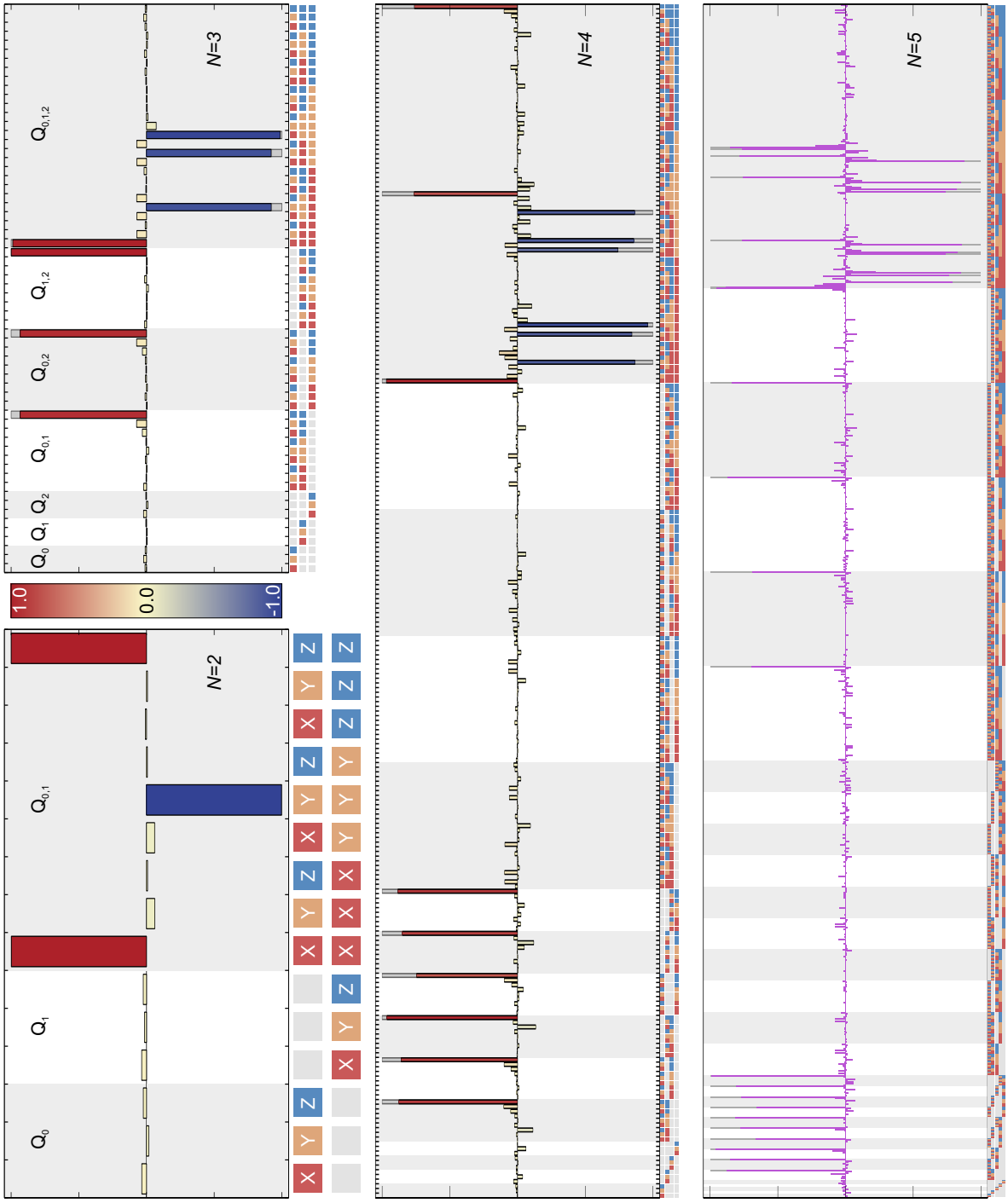


Figure B.10: **Pauli operator representation for the $N = 2$ Bell state and the $N = 3, 4$ and 5 GHZ states.** The bars show expectation values of combinations of Pauli operators, ideal in grey, experimental values in colour. Apart from the trivial even number Z-correlations, the data show significant elements only for the largest correlations.

Appendix C

Supplementary Information for Chapter 5: "Optimal quantum control using randomized benchmarking"

C.1 Scaling of the sensitivity of ORBIT with error

Here, we derive the sensitivity of the sequence fidelity to gate error, and show that the sensitivity to fractional error is constant – hence ORBIT can in principle scale to arbitrarily small errors.

The sequence fidelity decays with m following $F = Ap^m + B$. For the single-qubit case: $p = 1 - 2r$, with r the error per Clifford. The variation in sequence fidelity with gate error is

then

$$dF/dr = -2Am(1-2r)^{m-1}. \tag{C.1}$$

The optimal value of m to operate ORBIT is at the characteristic decay of the sequence fidelity

$$m' = -1/\ln(1-2r) \tag{C.2}$$

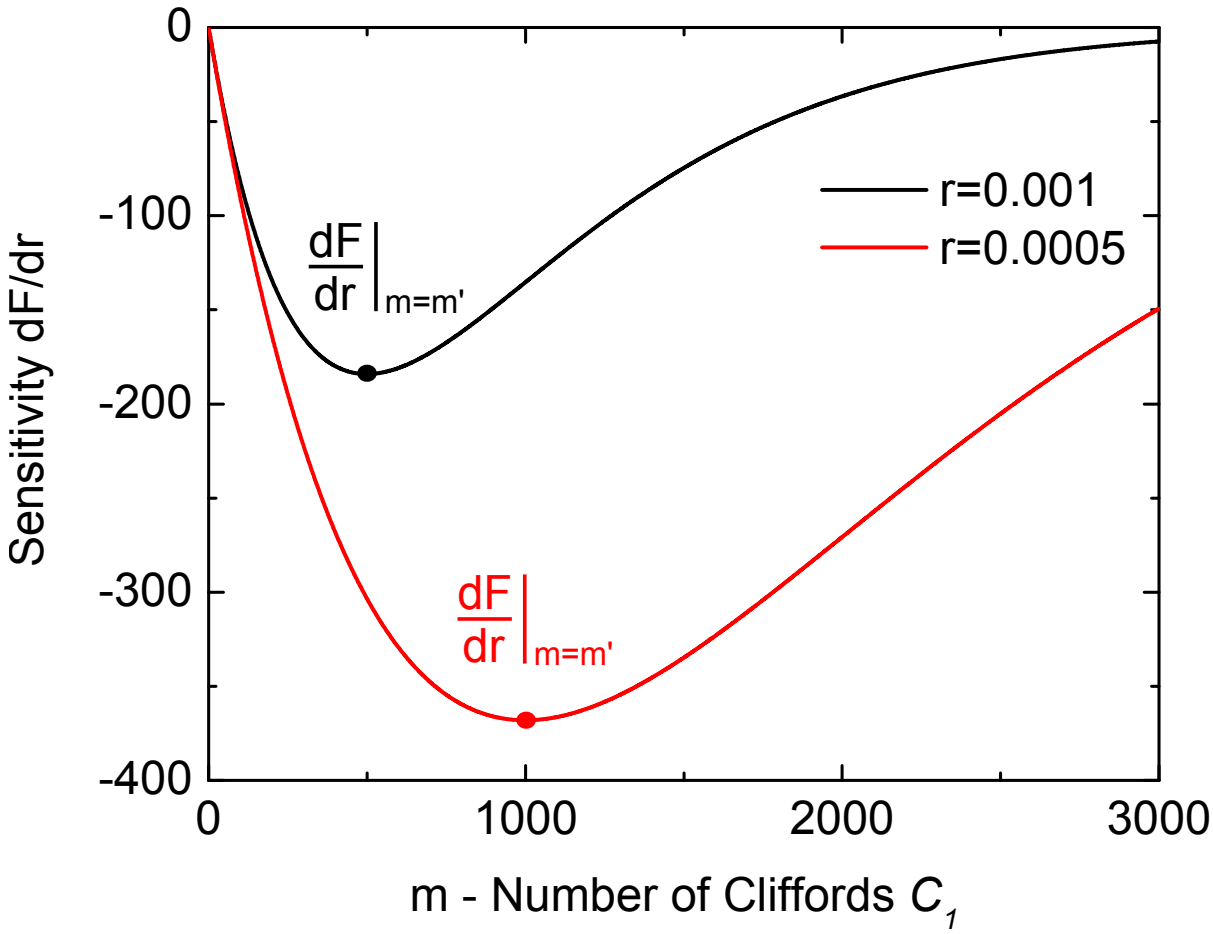


Figure C.1: The sensitivity of sequence fidelity to gate error (Eq. C.1) for error per Clifford $r = 0.001$ and $r = 0.0005$. We take the ideal value $A = 0.5$ for the single-qubit case. Dots represent the optimal value of $m = m'$ using Eq. C.2 and Eq. C.3.

(this becomes clear when expressing the sequence fidelity as $F = A\exp(-m/m') + B$). To quantify the scaling of the sensitivity with gate error, we evaluate the sensitivity at m', where dF/dr is maximal.

$$\left.\frac{dF}{dr}\right|_{m=m'} = \frac{2A}{e(1-2r)\ln(1-2r)} \approx -\frac{A}{er}, \tag{C.3}$$

with the right side an expansion for small r, keeping the lowest order term.

Importantly, the sensitivity to fractional error (dr/r) is constant,

$$S = \frac{dF}{dr/r} = r\left.\frac{dF}{dr}\right|_{m=m'} \approx -\frac{A}{e}. \tag{C.4}$$

This is a crucial result, as it implies that ORBIT scales to arbitrarily small error: the sensitivity is the same when improving the fidelity of a 99.0% gate to 99.9%, or a 99.99% gate to 99.999%; only the choice for m is different.

As an example, Eq. C.1 is plotted in Fig. C.1 for two cases: $r = 0.001$ and 0.0005 ($A = 0.5$). These cases reach a maximum sensitivity at $m' = 500$ and $m' = 1000$ respectively. When halving the error the optimal m and sensitivity double, as expected. We note that the sensitivity is retained for a wide range of m around the optimum, therefore the choice of m need not be exact. This is useful for improving gates, as we generally operate at a fixed m, and changes in r will affect m'.

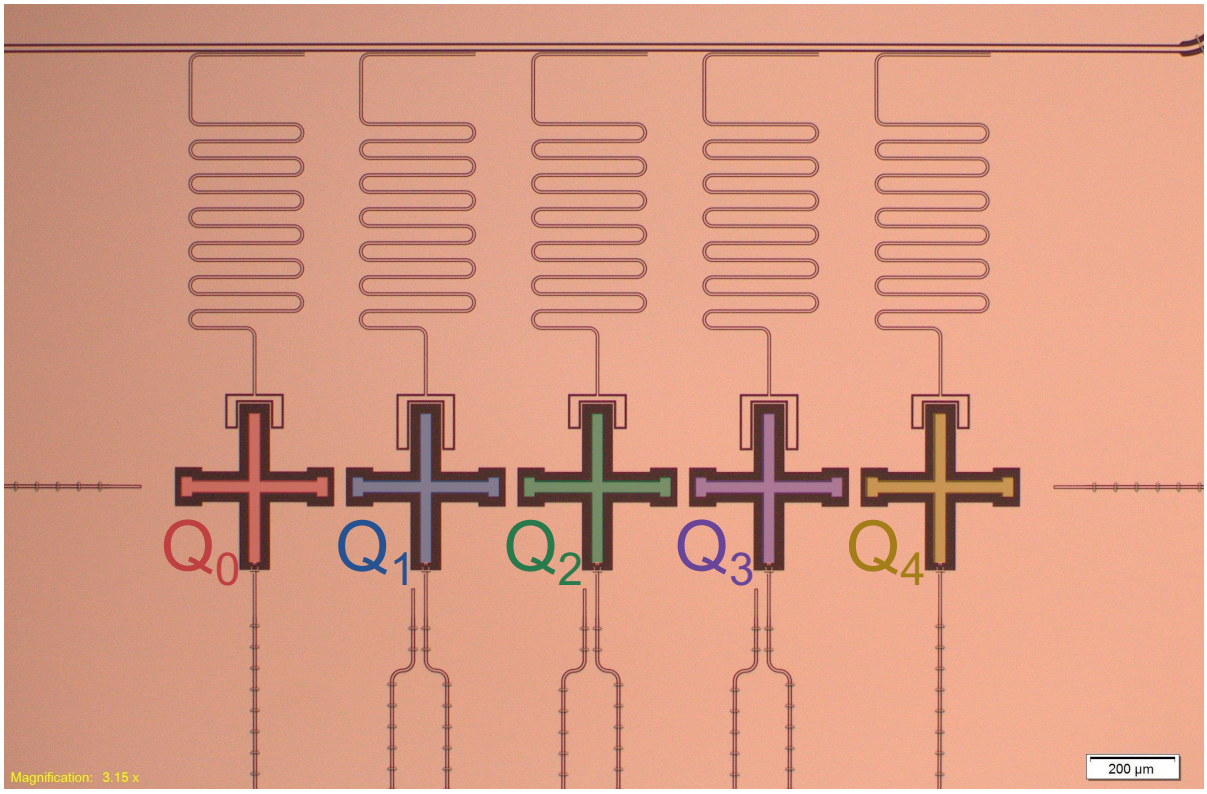


Figure C.2: Optical micrograph of the 5 Xmon transmon device. Qubits have individual XY and Z control with individual readout. Neighboring qubits have direct capacitive coupling of $g/2\pi = 30$ MHz.

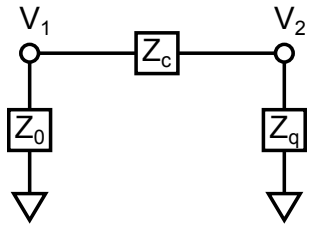


Figure C.3: Circuit model for qubit coupled to XY drive. Node 1 represents the XY drive line with impedance Z_0, which is capacitively coupled via Z_c to node 2 (the qubit) with characteristic impedance Z_q.

C.2 Device Parameters

qubits	Q_0	Q_1	Q_2	Q_3	Q_4
f_{10} (GHz)	5.805	5.238	5.780	5.060	5.696
nonlinearity (GHz)	-0.217	-0.226	-0.214	-0.212	-0.223

Additional information, including coherence times and fabrication details can be found in the Supplementary Information of Ref [**?**].

C.3 Experimental Setup

The wiring diagram and circuit components are shown in Fig. C.4, reproduced from Ref [**?**].

C.3.1 Electronics noise and drift

We find that eletronics noise have a negligible effect on qubit control. We measure an output noise temperature of $5 \cdot 10^3$ K from the room temperature electronics for the XY qubit control. There are two mechanisms that mitigate this noise: in-line attenuation and on-chip isolation. We have 40 dB in-line attenuation, and approximately 10 dB from cabling. The

on-chip isolation is set by the coupling capacitance from the XY drive line to the qubit, which is approximately 60 aF [**?**]. We model the circuit as a 50 Ω line capacitively coupled to a qubit of characteristic impedance Z_Q, as in Fig. C.3. The isolation Λ is

$$\Lambda = \frac{V_1^2/Z_0}{V_2^2/Z_q} \tag{C.5}$$

$$= \frac{Z_0 Z_q}{(Z_c + Z_q)^2} \tag{C.6}$$

Using $Z_0 = 50\,\Omega$, $Z_q = 300\,\Omega$, $Z_c = 1/i\omega C$, $C = 60$ aF and $\omega = 2\pi \cdot 6$ GHz and inserting into Eq. C.6 we get $\Lambda = 71$ dB isolation. This gives us a total of 121 dB attenuation, making the control electronics noise a negligible effect compared to the temperature of the environment.

We also must consider the noise coming from the Z control board. This will cause a jitter in the qubit frequency and manifest as dephasing. We find no difference in dephasing times when the Z control board is connected to the qubit or not, indicating this is not a limiting mechanism.

Lastly, we find that fidelities of the single- and two-qubit gates remain stable over the course of many hours from calibration, indicating that drifts in optimal parameters are small.

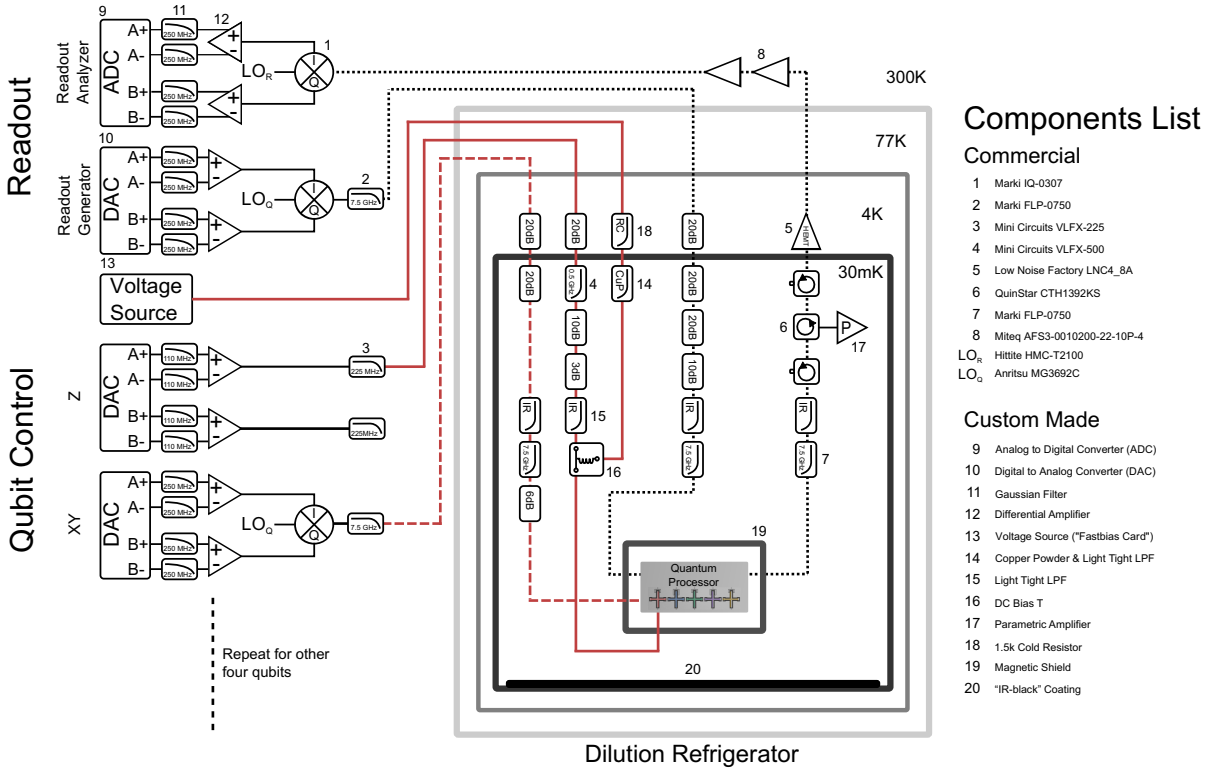


Figure C.4: (This figure reproduced from the Supplementary Information of Ref [**?**]) Electronics and Control Wiring. Diagram detailing all of the control electronics, control wiring, and filtering for the experimental setup. Each qubit uses one digital to analog converter (DAC) channel for each of the X, Y, and Z rotations. Additionally, we use a DC bias tee to connect a voltage source to each qubit frequency control line to give a static frequency offset. All five qubits are read out using frequency-domain multiplexing on a single measurement line. The readout DAC generates five measurement tones at the distinct frequencies corresponding to each qubit's readout resonator. The signal is amplified by a wideband parametric amplifier [**?**], a high electron mobility transistor (HEMT), and room temperature amplifiers before demodulation and state discrimination by the analog to digital converter (ADC). All control wires go through various stages of attenuation and filtering to prevent unwanted signals from disturbing the quantum processor. Two local oscillators (LO_Q) are used for qubit XY control, at 4.5 and 5.6 GHz. The readout LO_R is at 6.76 GHz. All LO, DAC, and ADC electronics are locked to a 10 MHz SRS FS725 rubidium frequency standard.

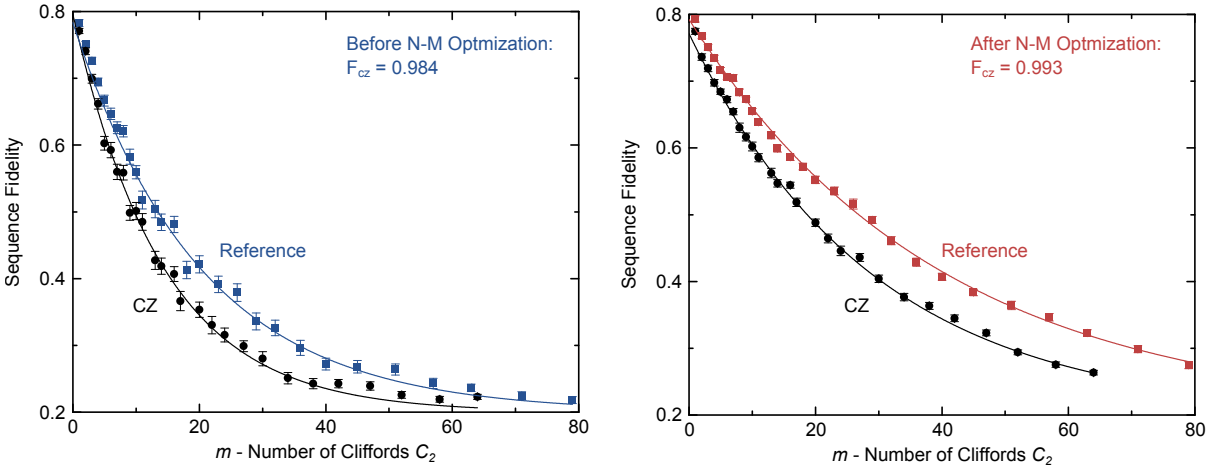


Figure C.5: Two-qubit randomized benchmarking data for Fig. 2 in the main text, showing the decay of the sequence fidelity of the reference and when interleaved with the CZ gate ($k = 50$). (a) Before Nelder-Mead optimization. Reference error: $r_{\mathrm{ref}} = 0.0361$, interleaved error: $r_{\mathrm{ref+CZ}} = 0.0511$, extracted CZ error: $r_{\mathrm{CZ}} = 0.0157$. (b) After Nelder-Mead optimization. Reference error: $r_{\mathrm{ref}} = 0.0188$, interleaved error: $r_{\mathrm{ref+CZ}} = 0.0254$, extracted CZ error: $r_{\mathrm{CZ}} = 0.0068$.

C.4 CZ gate fidelity before and after Nelder-Mead optimization

The reference and interleaved randomized benchmarking data for the CZ gate, for Fig. 2 in the main text, are shown in Fig. C.5. Figure C.5a is before improvement, Fig. C.5b after. The extracted fidelity of 0.993 is slightly lower than the extracted fidelity of 0.994 for the same pair of qubits in Ref. [**?**]. We attribute this to a small increase in the dephasing rate after thermally cycling the sample between experiments.

As a self-consistency check, we can calculate the expected error per Clifford using the derivation in Ref. [**?**]. We assume that gate errors are small and uncorrelated, such that adding errors is a good approximation. The expected error per Clifford is $r_{\mathrm{ref,predicted}} = 8.25\, r_{\mathrm{SQ}} +$

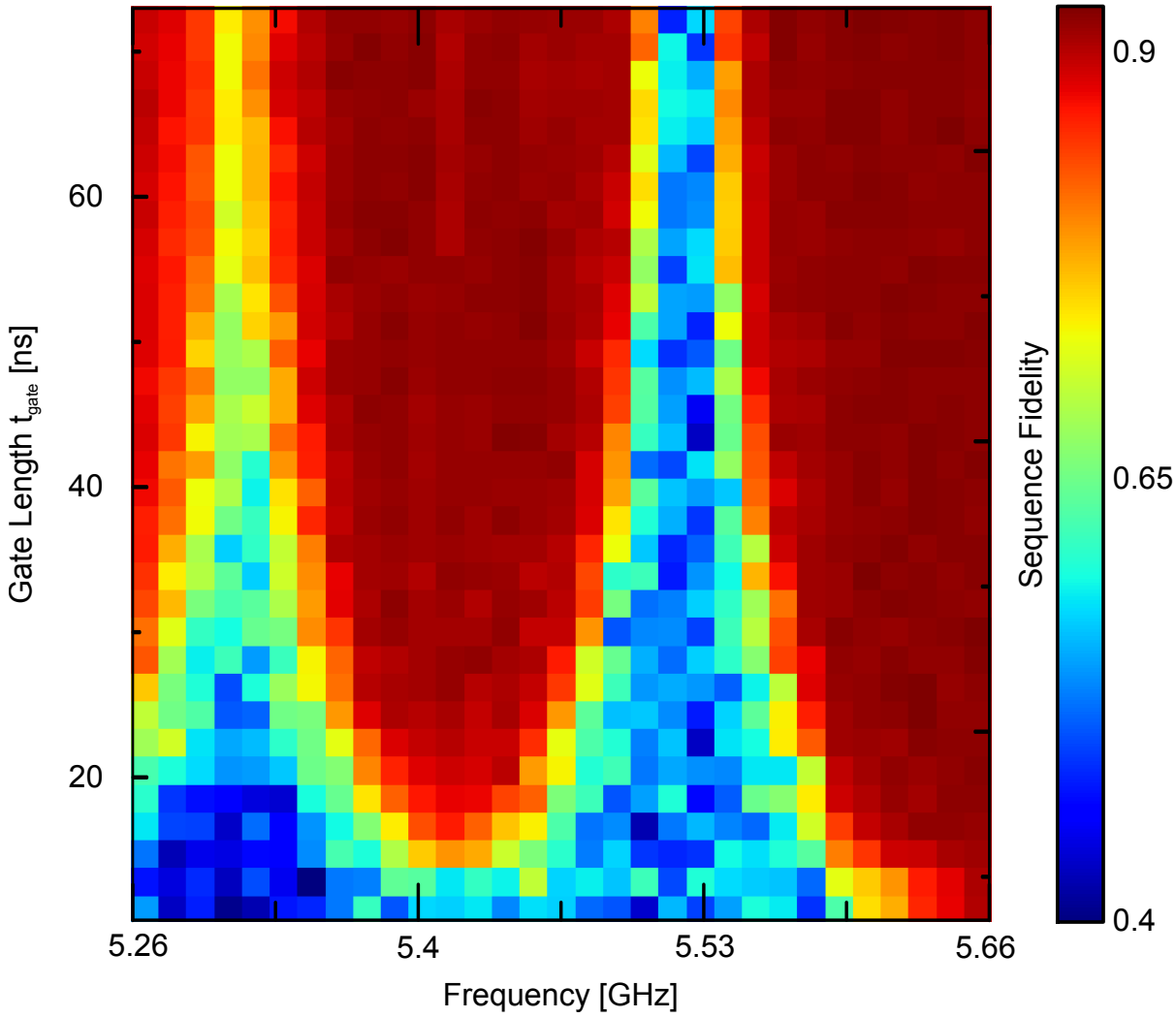


Figure C.6: Sequence fidelity data for Fig. 4 in the main text at $m = 35$ ($k = 20$).

$1.5\, r_{\mathrm{CZ}}$ with r_{SQ} the average single-qubit gate error and r_{CZ} the CZ gate error. Assuming $r_{\mathrm{SQ}} = 0.001$, we compute $r_{\mathrm{ref,predicted,before}} = 0.0318$ and $r_{\mathrm{ref,predicted,after}} = 0.0185$ which are close to the experimental values of $r_{\mathrm{ref,before}} = 0.0361$ and $r_{\mathrm{ref,after}} = 0.0188$.

C.5 Control crosstalk data

The sequence fidelity data, for Fig. 4 in the main text, are shown in Fig. C.6.

Appendix D

Supplementary information for Chapter 7: "Overcoming non-markovian noise in quantum systems: How mediocre clocks make good qubits"

D.1 Theory: relation of RB error to $\langle \phi^2 \rangle$

In order to determine the effect of various dephasing mechanisms on an RB sequence, we first consider the following simplified model: a single qubit begins in $|\psi_0\rangle = |0\rangle$, then a randomly chosen perfect Clifford rotation C_1 is applied, and then a phase $\phi_{g,n}$ is accumulated by application of a Z rotation to simulate phase noise. The random Clifford and noise gate pair

are repeated N times, after which the single Clifford C_r that is the inverse of all the previous Cliffords is applied to rotate back to (nearly) $|0\rangle$ and we measure the probability of error, $P_{err} = |\langle 1|\psi_N\rangle|^2$.

The value of $\phi_{g,n}$ depends on the dephasing model employed. For example, for static dephasing (e.g., a frequency offset), it is constant: $\phi_{g,n} = \phi_{g,st}$. For white noise, $\phi_{g,n}$ is randomly sampled from a symmetric Gaussian distribution. In general, $\phi_{g,n}$ is arbitrary, but we assume $|\phi_{g,n}| \ll 1$. The average square of $\phi_{g,n}$ is denoted $\langle \phi_g^2 \rangle$.

We now consider the "error angle", $\Delta\phi$, the angular separation of $|\psi_N\rangle$ from $|0\rangle$ in the Bloch sphere picture of a single qubit, noting that $P_{err} = \langle (\Delta\phi/2)^2 \rangle$, assuming $|\Delta\phi| \ll 1$. Because $|\phi_{g,n}| \ll 1$ and N is not too large, after each rotation $|\psi\rangle$ is close to one of the six axes ($\pm X, \pm Y, \pm Z$), and the angular distance from the axis is $\Delta\phi$. There is a 1/3 chance that the qubit is near the pole (i.e. Z axis) and then the rotation $\phi_{g,n}$ does not change $\Delta\phi$, while with 2/3 probability the qubit is near the equator and $\Delta\phi$ is changed.

For any dephasing model, it is straightforward to see that the evolution of $\Delta\phi$ is essentially a random walk in two dimensions, and that

$$\langle (\Delta\phi)^2 \rangle = \frac{2}{3} N \langle \phi_g^2 \rangle, \tag{D.1}$$

assuming $N\langle \phi_g^2 \rangle \ll 1$. The RB error is then

$$P_{err} = \langle (\Delta\phi/2)^2 \rangle = \frac{1}{6} N \langle \phi_g^2 \rangle. \tag{D.2}$$

It might be expected that in the static dephasing case there can be some sort of echoing effect; for example, if a Clifford takes $|\psi\rangle$ to the +Y axis and it is rotated by $\phi_{g,st}$, then if

the next Clifford is an X rotation, putting $|\psi\rangle$ near the -Y axis, the following rotation also by $\phi_{g,st}$ will cancel the previous noise rotation. However, when the full set of Clifford rotations is used, there are four rotations that take $|\psi\rangle$ near the -Y axis, and each orients the previous $\Delta\phi$ in a different direction relative to the axis, resulting in equal probability of canceling the previous rotation, doubling it, or moving in one of the two perpendicular directions. The noise accumulated between rotations is therefore uncorrelated with previous or future noise; the Clifford set is error depolarizing. Therefore, Eqs. (D.1) and (D.2) hold regardless of the noise model.

This simplified model has been confirmed with simulation, for both a static and an uncorrelated noise model with $\phi_{g,n} = \pm\phi_g$.

This implies that RB is an effective way to measure dephasing, if the sequence error occuring between the gates is attributable to dephasing. This can be done easily by comparing the sequence fidelity of an RB sequence with interleaved idling time to that of a reference RB sequence, effectively subtracting out errors due to the Clifford gates themselves–in other words, measuring the fidelity of an idle using interleaved RB, as in [**?**]. We can therefore measure the dephasing that takes place during an idle, and by varying the length τ of an idle, measure dephasing as a function of time, $\langle\phi^2(\tau)\rangle$ (for brevity we removed the subscript g). With $r_{I(\tau)}$ being the error rate (i.e. error per gate) of an idle, we thus arrive at Eq. (1) in the main paper:

$$P_{err}/N = r_{I(\tau)} = \frac{1}{6}\langle\phi^2(\tau)\rangle. \tag{D.3}$$

For completeness, we also mention here the effect of energy relaxation (T_1 decay) on the

fidelity of RB sequences. After each Clifford, the qubit state $|\psi\rangle$ is near the equator of the Bloch sphere with probability 2/3. In this case the probability of the energy relaxation event is $\tau/2T_1$ (we assume $\tau \ll T_1$); such an event moves $|\psi\rangle$ by approximately the angle $\pi/2$ on the Bloch sphere, thus leading to the error probability 1/2 at the end of the RB sequence. The corresponding contribution to the RB error per gate is $(2/3) \times (\tau/2T_1) \times (1/2) = \tau/6T_1$. With probability 1/6 the qubit state after a Clifford is close to the North pole (state $|0\rangle$); then there is no energy relaxation. Finally, with probability 1/6 the qubit state is close to the South pole $|1\rangle$; then the probability of the energy relaxation event is τ/T_1, which moves the state by approximately the angle π, thus almost certainly leading to the RB error. The corresponding contribution to the RB error per gate is $(1/6) \times (\tau/T_1) \times 1 = \tau/6T_1$. Adding together the two contributions, we arrive at

$$P_{err}/N = \frac{\tau}{3T_1}. \tag{D.4}$$

Since T_1 can be measured independently, the effects of T_1 decay can be calculated and subtracted from the results obtained with RB, much as it can be subtracted from Ramsey visibility decays as well. In our experiment T_1 is relatively large, and therefore this correction is small.

D.2 Types of Phase Noise

We now discuss the form of $\langle\phi^2(\tau)\rangle$ for different sources of noise. For completeness, we also show the similar characteristic, $\langle\tilde{\phi}^2(\tau)\rangle$, for the echo sequence of duration τ (with π pulse at $\tau/2$). Most of results discussed here were presented earlier, e.g., in Refs. [**?**][**?**][**?**].

The average values $\langle\phi^2(\tau)\rangle$ and $\langle\tilde{\phi}^2(\tau)\rangle$ for the idle and echo sequence, respectively, can

be calculated via the spectral density $S(\omega)$ of the qubit frequency fluctuation,

$$\langle \phi^2(\tau) \rangle = \tau^2 \int_0^\infty S(\omega) \left(\frac{\sin(\omega\tau/2)}{\omega\tau/2} \right)^2 \frac{d\omega}{2\pi}, \tag{D.5}$$

$$\langle \tilde{\phi}^2(\tau) \rangle = \tau^2 \int_0^\infty S(\omega) \frac{\sin^4(\omega\tau/4)}{(\omega\tau/4)^2} \frac{d\omega}{2\pi}, \tag{D.6}$$

where $S(\omega)$ is single-sided and the average frequency fluctuation is assumed to be zero.

For the white noise with a flat spectral density, $S(\omega) = S_0$, we find

$$\langle \phi^2_{\text{white}}(\tau) \rangle = \langle \tilde{\phi}^2_{\text{white}}(\tau) \rangle = \frac{S_0}{2}\tau = 2\frac{\tau}{T_{\phi 1}}, \tag{D.7}$$

where $T_{\phi 1} = 4/S_0$ is the dephasing time due to white noise. Note that the factor of 2 in the last expression cancels when the corresponding visibility of a Ramsey or echo sequence, $\exp(-\tau/T_{\phi 1})$, is calculated.

For noise that is correlated over very long times (very slowly fluctuating qubit frequency), $S(\omega) = 4\pi\sigma_{qb}^2\delta(\omega)$, where σ_{qb} is the standard deviation of the qubit frequency $2\pi f_{10}$. In this case

$$\langle \phi^2_{\text{corr}}(\tau) \rangle = \sigma_{qb}^2\tau^2 = 2\left(\frac{\tau}{T_{\phi 2}}\right)^2, \;\; \langle \tilde{\phi}^2_{\text{corr}}(\tau) \rangle = 0, \tag{D.8}$$

where $T_{\phi 2} = \sqrt{2}/\sigma_{qb}$ is the Ramsey dephasing timescale due to such correlated noise. Obviously, in this case there is no dephasing in the echo sequence visibility.

For $1/f$ noise let us use $S(\omega) = \dfrac{S_{1/f}}{\omega/2\pi}$, then [**? ?**]

$$\langle \phi^2_{1/f}(\tau) \rangle = S_{1/f}\,\tau^2 \ln\frac{0.4007}{f_c\tau}, \tag{D.9}$$

$$\langle \tilde{\phi}^2_{1/f}(\tau) \rangle = S_{1/f}\,\tau^2 \ln 2, \tag{D.10}$$

where $f_c = \omega_c/2\pi$ is the low-frequency cutoff of the $1/f$ noise (e.g., the inverse of the total

duration of the experiment), which is introduced as the lower limit of integration in Eq. (D.5). Note that in Eq. (D.9) we assumed $f_c\tau \lesssim 0.2$. As the log part in Eq. (D.9) varies slowly, typically it is ignored and $1/f$ noise for $\langle\phi^2(\tau)\rangle$ is treated with Eq. (D.8). Note that the factors in Eq. (D.9) and (D.10) are different, resulting in different effective dephasing times $T_{\phi 2}$ for the Ramsey and echo sequences.

Finally, let us consider a telegraph noise, for which the qubit frequency $2\pi f_{10}$ switches between two values separated by $\Delta\omega_{qb}$, with up (down) switching rate of $\Gamma_\uparrow$ ($\Gamma_\downarrow$). In this case

$$S(\omega) = \frac{4(\Delta\omega_{qb})^2\Gamma_\uparrow\Gamma_\downarrow}{\Gamma_\Sigma(\omega^2+\Gamma_\Sigma^2)},\ \ \Gamma_\Sigma = \Gamma_\uparrow + \Gamma_\downarrow, \tag{D.11}$$

so using Eqs. (D.5) and (D.6) we obtain

$$\langle\phi^2_{\text{tel}}(\tau)\rangle = 2\frac{(\Delta\omega_{qb})^2}{\Gamma_\Sigma}\frac{\Gamma_\uparrow\Gamma_\downarrow}{\Gamma_\Sigma^2}\left(\tau - \frac{1-e^{-\Gamma_\Sigma\tau}}{\Gamma_\Sigma}\right), \tag{D.12}$$

$$\langle\tilde{\phi}^2_{\text{tel}}(\tau)\rangle = 2\frac{(\Delta\omega_{qb})^2}{\Gamma_\Sigma}\frac{\Gamma_\uparrow\Gamma_\downarrow}{\Gamma_\Sigma^2}\left(\tau - \frac{3+e^{-\Gamma_\Sigma\tau}-4e^{-\Gamma_\Sigma\tau/2}}{\Gamma_\Sigma}\right). \tag{D.13}$$

Note that at short time, $\tau \ll \Gamma_\Sigma^{-1}$, the effect of the telegraph noise is similar to the effect of the correlated noise with $T_{\phi 2} = \sqrt{2}\,\Gamma_\Sigma/(\sqrt{\Gamma_\uparrow\Gamma_\downarrow}\,\Delta\omega_{qb})$, while at long time, $\tau \gg \Gamma_\Sigma^{-1}$ it is similar to the effect of white noise with $T_{\phi 1} = \Gamma_\Sigma^3/[\Gamma_\uparrow\Gamma_\downarrow(\Delta\omega_{qb})^2]$.

Defining the effective switching amplitude as $2\pi\Delta f_{10} = \Delta\omega_{qb}\sqrt{\Gamma_\uparrow\Gamma_\downarrow}/\Gamma_\Sigma$ and introducing notation $T_{\text{sw}} = 1/\Gamma_\Sigma$, we can rewrite Eq. (D.12) as

$$\langle\phi^2_{\text{tel}}(\tau)\rangle = 2(2\pi\Delta f_{10})^2 T_{\text{sw}}[\tau - T_{\text{sw}}(1-e^{-\tau/T_{\text{sw}}})], \tag{D.14}$$

which is Eq. (2) from the main paper. Note that the telegraph noise in not Gaussian. Therefore, while the obtained equations can be used to find the RB error per gate, they cannot,

strictly speaking, be used to find the visibility of the standard Ramsey and echo sequences. Nevertheless, they can be used approximately if $|\Delta\omega_{qb}|/\min(\Gamma_\uparrow, \Gamma_\downarrow) \ll 1$, because at short time the accumulated phase shift is small and the assumption of Gaussianity is not needed, while at longer time, when the phase becomes comparable to 1, the probability distribution for the phase becomes Gaussian due to a large number of switching events.

D.3 T_1, Ramsey, and spin echo fits

The T_1 data are fit to a simple exponential, $P_1(t) = A\exp(t/T_1)+B$, and we find $T_1 = 26.7\,\mu$s. The Ramsey and spin scho envelopes are each fit to a noise model that includes white and correlated components,

$$V(t) = A\exp\left[-t/T_{\phi 1} - (t/T_{\phi 2})^2\right] + B, \tag{D.15}$$

where $V(t)$ is the Ramsey/echo visibility, t is the length of the idle as shown in Fig. 1 of the main text, $T_{\phi 1}$ is the white noise dephasing timescale, $T_{\phi 2}$ is the correlated noise dephasing timescale, and A and B are the result of state preparation and measurement errors. The fit parameters are given below. Note that each of the fits includes the full range of data, from $0 < t < 5.0\,\mu$s for Ramsey and $0 < t < 12.0\,\mu$s for Echo.

Sequence	$T_{\phi 1}$ (μs)	$T_{\phi 2}$ (μs)	A	B
Ramsey	6.8	2.8	0.88	0.015
Spin Echo	15.1	7.5	0.88	0.021

D.4 Flux noise

Flux noise on this device, plotted in Fig. D.1, has been measured over the frequency range $10^{-4} < f < 1$ Hz, using the Ramsey Tomography Oscilloscope (RTO) protocol of repeated frequency measurements as described in [**?**]. Four measurements were made on this device

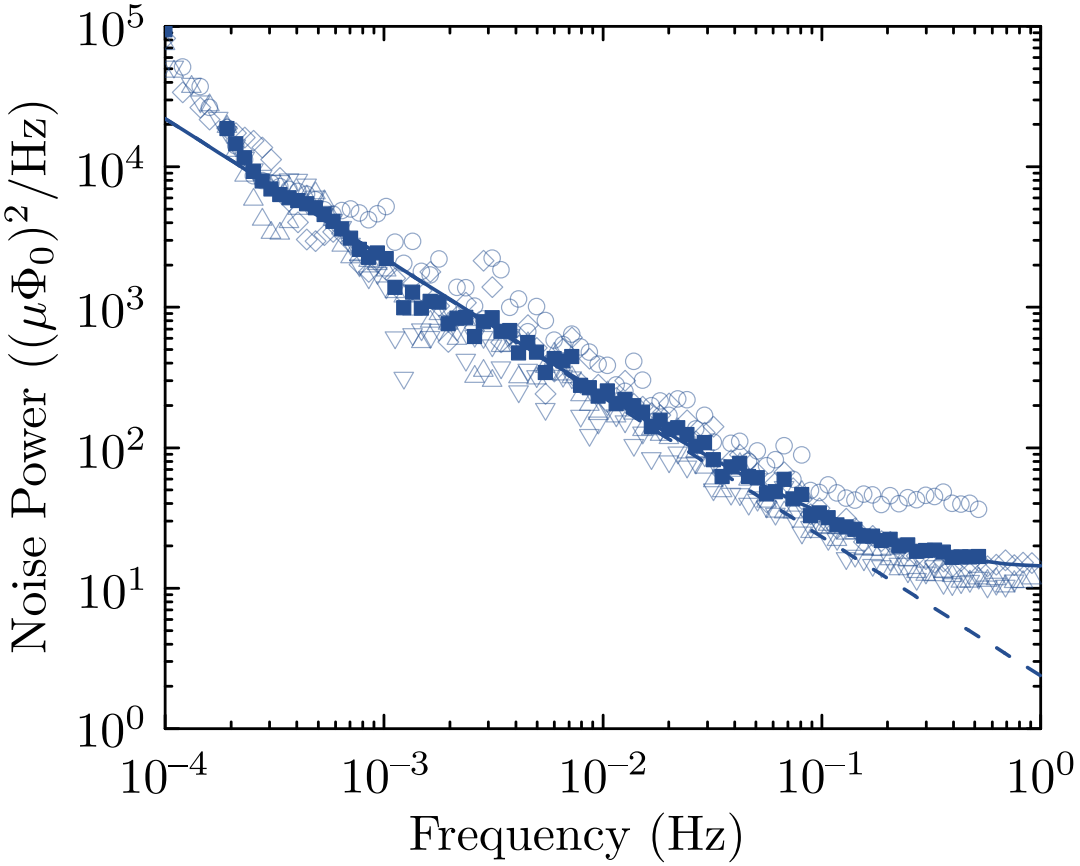


Figure D.1: Flux noise as measured with RTO [**?**]

(open markers), at three different operating points, and then each measurement was binned in log-space, and the binned measurements averaged together (closed squares). This average is fit (solid line) to an aliased $1/f$ and white noise model, given by

$$S_\phi(f) = S_\phi^*/f^\alpha + S_\phi^*/(2f_n - f)^\alpha + S_{\text{white}}, \tag{D.16}$$

where $S_\phi(f)$ is the flux noise power, expressed in $(\mu\Phi_0)^2/\text{Hz}$, f is the noise frequency, α is the slope of the noise (1 for pure $1/f$ noise), S_ϕ^* is the flux noise power at 1 Hz, $f_n = 1$ Hz is the Nyquist frequency of the measurement, and S_{white} is the white noise floor. From the fit we extract $S_\phi^* = 2.4\,(\mu\Phi_0)^2$, $\alpha = 0.99$, and $S_{\text{white}} = 9.7\,(\mu\Phi_0)^2/\text{Hz}$. We attribute the white noise to state preparation and measurement error. The dashed line shows the $1/f$ fit extended to 1 Hz, where the value of the y-intercept is S_ϕ^*.

To plot the inferred flux noise contribution in Fig. 2 in the main paper and Fig. D.2 below, we use Eq. (D.9), with $S_{1/f} = \partial f/\partial\phi \cdot S_\phi^*$ taken from the measurements above, and $f_c = 10$ min, the length of the experiment. The value of the log factor of Eq. (D.9) varies from 13 to 7 for $1 < \tau < 450$ ns.

D.5 RB Ramsey across the qubit spectrum

Figure D.2 shows RB Ramsey measurements at three additional qubit frequencies; the data for the 4.9 GHz operating point are the same as in Fig. 2 of the main paper. The inset shows the frequency-flux relation for this qubit, with the four operating points denoted by symbols; $df/d\phi$ changes by a factor of 2.7 between the operating points, to explore different susceptibilities to flux noise. The remaining three datasets are fit to a noise model incorporating telegraph and

white noise; that is,

$$r_{I(\tau)} = \frac{1}{6}\left(\langle\phi^2_{\text{tel}}(\tau)\rangle + \langle\phi^2_{\text{white}}(\tau)\rangle\right) \tag{D.17}$$

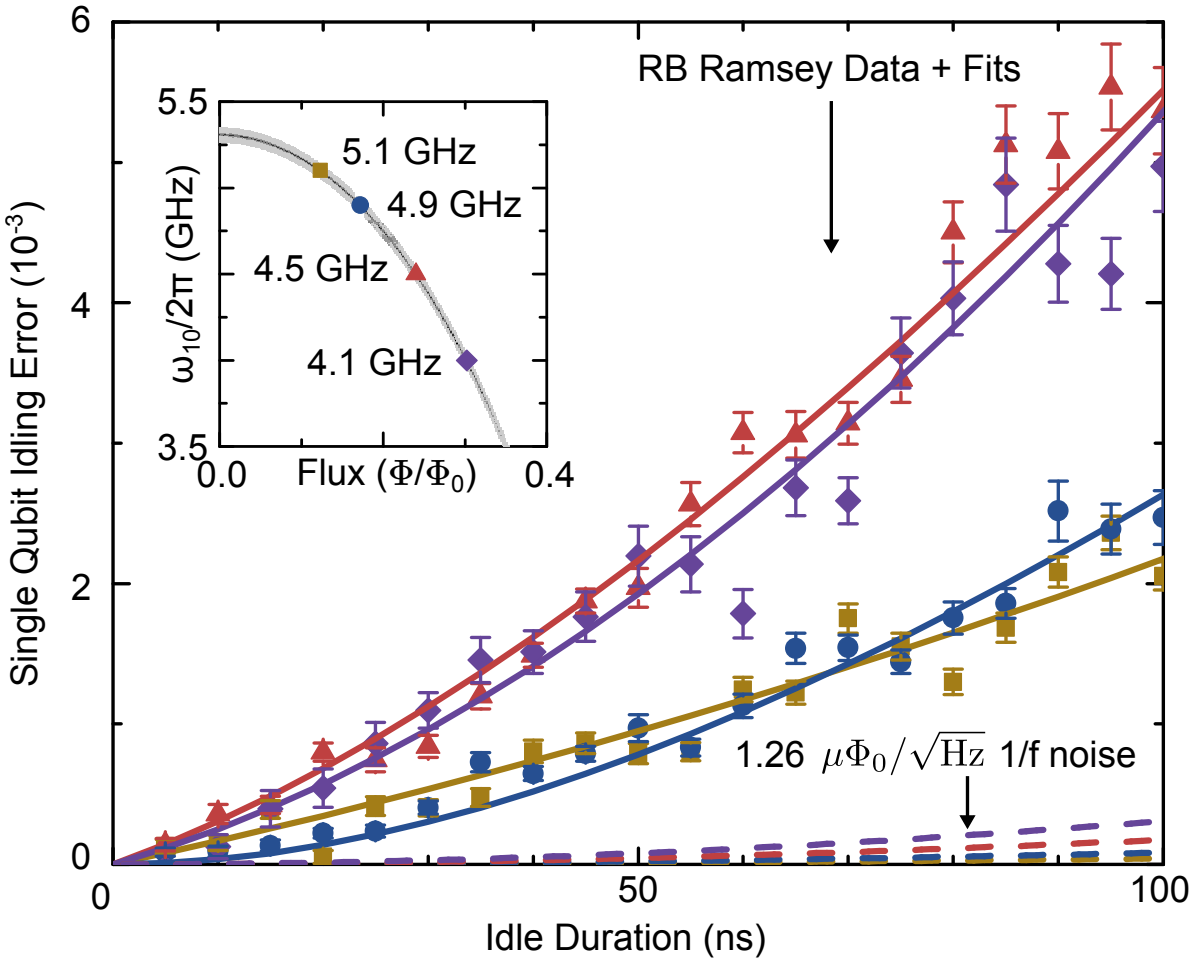


Figure D.2: (color online) RB Ramsey idling error *vs.* duration, for various frequencies; T_1 effects have been subtracted according to Eq. (D.4). The dashed lines denote the inferred contribution from $1/f$ flux noise at the four different operating points. The inset shows frequency spectroscopy *vs.* applied flux, following the expected dependence [**?**]; the four operating points are shown.

[see Eqs. (D.7) and (D.14)]. We show the fit parameters here.

f_{10} (GHz)	$df/d\Phi$ (GHz/Φ_0)	T_1 (μs)	$T_{\phi 1}$ (μs)	T_{sw} (ns)	Δf_{10} (kHz)
5.1	3.39	30.6	20.6	182000	184
4.9	4.81	26.7	-	84	479
4.5	6.95	31.3	12.4	98	484
4.0	9.23	36.2	15.5	263	469

We note that at the highest qubit frequency, the large T_{sw} indicates that the telegraph noise model is not needed here and can be replaced by the correlated noise model with $T_{\phi 2} = \sqrt{2}/[2\pi\Delta f_{10}] = 1.2\,\mu$s. The Ramsey data for this frequency, fit to Eq. (D.15), give $T_{\phi 1} = 10.7\,\mu$s and $T_{\phi 2} = 3.6\,\mu$s, which indicates that even though the telegraph dephasing source is not present at this operating point, the dephasing magnitude measured by Ramsey still does not match that found with RB.

Despite tuning the flux Φ/Φ_0 over most of its range, we find that $1/f$ noise does not contribute appreciably to gate errors. For typical gates of length 20ns, idle fidelities greater than 0.999 are seen over the frequency range, demonstrating that tunable qubits can achieve high fidelity even when biased significantly away from the flux-insensitive operating point.

D.6 Charge Noise

To calculate the expected frequency fluctuation due to charge noise, we use Eq. (2.5) from [**?**]

$$\epsilon_m \simeq (-1)^m E_C \frac{2^{4m+5}}{m!}\sqrt{\frac{2}{\pi}}\left(\frac{E_J}{2E_C}\right)^{\frac{m}{2}+\frac{3}{4}} e^{-\sqrt{8E_J/E_C}}, \tag{D.18}$$

where ϵ_m is the charge dispersion for energy level m, and E_J and E_C are the Josephson energy and charging energy, respectively, of the qubit. Note that we can also write $E_J/E_C \approx (\omega_{01}/\eta - 1)^2/8$ (following from Eq. (2.11)), where $\omega_{01}/2\pi$ is the qubit frequency and the qubit anharmonicity $f_{12} - f_{01} = \eta/2\pi = -215$ MHz. We then calculate ϵ_1 for the two ends of the qubit spectrum; we find $\epsilon_1(\omega_{01}/2\pi = 6\text{ GHz}) = 3.6$ Hz and $\epsilon_1(\omega_{01}/2\pi = 4\text{ GHz}) = 14.4$ kHz, both of which are far below the measured charge noise fluctuation frequency of ≈ 500 kHz. We also note the qubits used in Ref. [**?**] have charge noise fluctuations of the same order as the telegraph noise measured here, but charge noise of that magnitude is expected, as explained by the different parameter range of those qubits: $\omega_{01}/2\pi = 4.387$ GHz and $\eta/2\pi = -334$ MHz, giving $\epsilon_1 \approx 2$ MHz.

D.7 Calculation of Ω_{ZZ}

Two capacitively coupled qubits have an XX-type coupling of the form $g(|01\rangle\langle 10| + |10\rangle\langle 01|)$, where the coupling constant g is half the swap rate between the qubits. The interaction between the higher levels, $\sqrt{2}g(|11\rangle\langle 20| + |02\rangle\langle 11|) + \sqrt{2}g(|11\rangle\langle 20| + |02\rangle\langle 11|)$, results in a repulsion of the $|11\rangle$ level from the $|02\rangle$ and $|20\rangle$ levels; this energy shift in the $|11\rangle$ level produces a ZZ-type interaction between the qubits. In the far-detuned limit, neglecting the XX-coupling,

the two-qubit Hamiltonian becomes

$$H = \omega_1|10\rangle\langle 10| + \omega_2|01\rangle\langle 01| + (\omega_1 + \omega_2 + \Omega_{ZZ})\,|11\rangle\langle 11|, \tag{D.19}$$

$$\Omega_{ZZ} = \frac{2g^2}{\Delta - \eta_2} + \frac{2g^2}{-\Delta - \eta_1}, \tag{D.20}$$

where ω_n and η_n are the qubit frequencies and nonlinearities, respectively, and $\Delta = \omega_1 - \omega_2$. In our system, $\eta_1 = \eta_2 \equiv \eta$, giving

$$\Omega_{ZZ} = \frac{4g^2\eta}{\Delta^2 - \eta^2}. \tag{D.21}$$

When both qubits are simultaneously performing an RB sequence, phase error ϕ per idle gate in qubit A is

$$\phi = \pm\frac{\Omega_{ZZ}}{2}t_{\text{gate}} \tag{D.22}$$

where t_{gate} is the idle gate duration, and the frequency shift $\pm\Omega_{ZZ}/2$ assumes centering the qubit frequency. This gives $\langle\phi^2\rangle = (\Omega_{ZZ}t_{\text{gate}})^2/4$, and since for RB the error per gate is $E = \langle\phi^2\rangle/6$ [see Eq. (D.3)], we arrive at Eq. (5) of the main paper for the error per gate due to the Ω_{ZZ} interaction,

$$E = \frac{\pi^2}{6}\left(\frac{\Omega_{ZZ}}{2\pi}t_{\text{gate}}\right)^2. \tag{D.23}$$

D.8 Fits to Gate Errors in Figure 4

For the data in the Fig. 4 of the main paper, the fits are made either to a simple linear model in the case of Markovian noise (the XX and YX cases) or to a quadratic and linear model in the case of non-Markovian noise (the I and Z cases). There is no offset in any fit.

Gate	Linear Term (10^{-6} error/ns)	Quadratic Term (10^{-6} error/ns^2)
I	17	0.22
XX	20	-
Z	24	0.18
YX	22	-

Note that the contribution from $T_1 = 26.7\,\mu$s to the linear portion of the error, given by Eq. D.4, is 9.3×10^{-6} error/ns, or roughly half of the error measured. The remainder is equivalent to a white noise dephasing with time constant $T_{\text{white}} \approx 30\,\mu$s, according to Eqs. D.3 and D.7. The quadratic terms correspond with $T_{\phi 2} \approx 1\,\mu$s.

Appendix E

Supplementary information for Chapter 8: "State preservation by repetetive error detection in a superconducting quantum circuit."

E.1 The classical repetition code

Suppose we wish to reliably store a single classical bit of information, 0 or 1, for a very long period of time. There are many ways we could attempt to do this. We could write the number on a piece of paper, or carve the number into a boulder, or even scratch the number into a diamond; however, all of these methods are error-prone. Paper burns, ink fades, rocks weather,

and diamonds can be stolen. There is no known physical method of truly permanently storing information for later retrieval. Any storage scheme will have some probability of failure per unit time, with the most likely failure mechanism in many schemes being negligent or malicious human activity. We shall quantify all of these failure mechanisms by a single number p, the probability of failure per unit time. Without loss of generality, we shall take failure to mean a bit-flip. We can always convert arbitrary errors into bit-flip errors, as lost or unrecognizable data can simply be replaced with a random 0 or 1.

The simplest method of increasing the reliability of information storage is to make multiple copies, and ensure these n copies are subject to errors that are as independent as possible. When using this generic method, we are using a classical repetition code. Multiple paper copies, for example, could be stored in multiple geographic locations. This won't prevent a planet-killing asteroid simultaneously destroying all data, or an organized multi-site human attack, however we shall assume these n-bit correlated errors are sufficiently rare to neglect. We shall also assume all m-bit, $1 < m < n$ correlated errors are sufficiently rare to neglect and focus on the probability p of each individual bit suffering a flip.

E.1.1 Trusted supervisor

Suppose we have a hypothetical perfectly reliable and trustworthy supervisor at our disposal. Once per unit time, this supervisor could check each bit of data. Each bit has independent probability p of having suffered a bit-flip. If p is small and n is large, most of the data is very likely to still be correct, and the supervisor can take a majority vote, and set the minority bits

to the majority value. Note that it is possible that a majority of sites will have suffered an error, and that after "correction" every site will contain the wrong value. The probability of a majority of sites suffering an error is

$$p_{\text{fail}} = \sum_{i=\lceil n/2 \rceil}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) p^i (1-p)^{n-i}. \tag{E.1}$$

Given the average number of errors is pn, and the standard deviation $\sigma = \sqrt{np(1-p)}$, we need $p < 1/2$ to ensure that the average number of errors is less than $n/2$. As n is increased, pn will then be an $O(\sqrt{n})$ increasing number of σ below $n/2$, implying exponential suppression of p_{fail}. This example shows how the classical repetition code, given independent errors and a trusted supervisor, can arbitrarily reliably store a single bit of information using only a simple majority vote per unit time. Note that to first order in p, $p_{\text{fail}} \sim p^{\lceil n/2 \rceil}$.

E.1.2 Secret data

Suppose now that we wish to keep the data secret. Instead of granting our supervisor permission to look at the data directly, we allow them only to ask if two given bits are the same or not. That is, we allow them to access the result of the exclusive-OR (XOR) operator $\oplus$, where $0 \oplus 0 = 0$, $0 \oplus 1 = 1$, $1 \oplus 0 = 1$, and $1 \oplus 1 = 0$. We shall conceptually arrange our n bits in a line, and focus on the XOR of neighboring bits.

When a single error occurs, away from the ends of the line of bits, the parities of two pairs of bits around the error become 1. Figure E.1 gives a detailed worked example showing how this simple case is handled. We represent the parity changes in the graph of Fig. E.1d as red

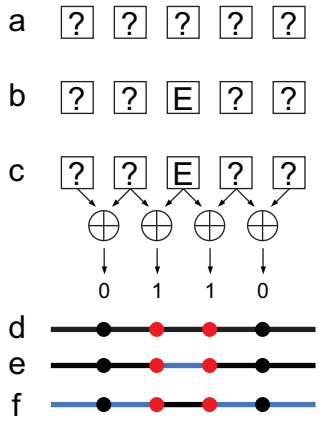


Figure E.1: **Processing perfect parity measurements. (a)** Five identical copies of an unknown bit represented by a question mark. **(b)** A bit-flip error on the middle bit, indicated by E. **(c)** Perfect measurement of the parity of neighboring bits would give all 0's in the absence of errors, but in this case two 1's are reported. **(d)** The string of parity measurements can be converted into a graph problem. Each vertex represents one parity measurement, and a red vertex is associated with each error. These 1's are located at the ends of chains of errors. Since errors are independent, error patterns containing fewer errors are more likely. **(e)** The simplest solution is a single error occurring between the red vertices, indicated by a blue line. **(f)** A second but less likely solution is the inverse of the above solution, consisting of two error chains and 4 errors.

colored vertices. The lines between the vertices represent how a single bit error event affects only its two neighboring vertices, except at the ends were the line connects only to an edge vertex. In Fig. E.1e we show the obvious decoding of these pair of red vertices into an error chain of a single blue line, which represents the bit-flip. However, the error can also be decoded into the 4 blue lines (error chains) in Fig. E.1f, which represents the inverse of the solution in Fig. E.1e. As an error chain with n lines has probability $O(p^n)$, assuming independent errors, the solution in Fig. E.1e with $O(p)$ is much more likely to occur than for Fig. E.1f with $O(p^4)$, and should thus be the choice for the decoding.

Figure E.2 gives a more complex example when there are multiple error events. To decode this example, we first show in Fig. E.2b the graph of the errors. To decode the multiple red

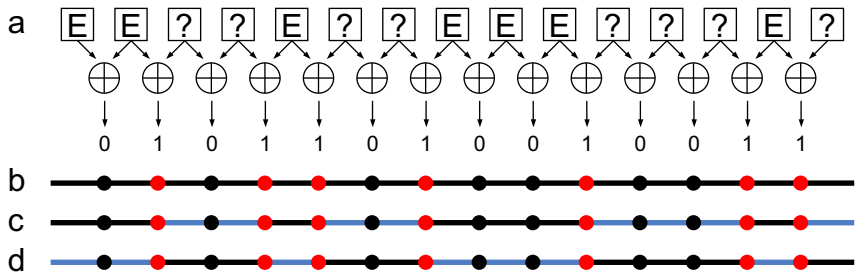


Figure E.2: **Processing perfect parity measurements, large example. (a)** 15 initially identical copies of an unknown bit suffer errors at unknown locations leading to a specific pattern of pairwise parity measurements. **(b)** Graph problem corresponding to the parity measurements. Our goal is to connect the colored vertices in pairs or to a graph boundary using the minimum total number of edges. **(c)** Non-optimal weight 8 graph solution. **(d)** Optimal weight 7 graph solution, i.e. contains the fewest errors. Note that after applying corrections corresponding to the blue edges, we will restore the data to its original state using the optimal graph solution, and the bit-inverse of the original state using the non-optimal graph solution. This is a generic property — after correction one will always obtain the original data or its perfect bit-inverse.

vertices into error events (blue lines), we need to form error chains. A simple algorithm to do this is: (I) From left to right, find the first red vertex and match it with another red vertex to its right. After matching a pair of vertices, continue matching pairs until reaching the right edge. (II) A second solution is the inverse of the above solution. These two solutions are shown in Fig. E.2c and d. As d has a shorter total length of the error chains, it is the most likely solution.

The errors are decoded properly if less than $\lceil n/2 \rceil$ errors occur in a single time interval. This means that only having access to parity information is just as powerful in decoding as being able to directly view the data, and we retain $p_{\text{fail}} \sim p^{\lceil n/2 \rceil}$. Indeed, the corrections suggested by matching perfect parity information are identical to the corrections suggested by taking a majority vote given actual data values.

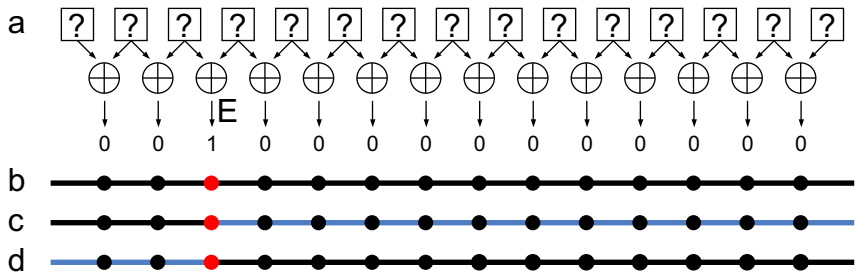


Figure E.3: **Incorrectly processing imperfect parity measurements. (a)** 15 initially identical copies of an unknown bit remain free of error, however a single parity measurement reports the wrong result. **(b)** The incorrect parity measurement leads to a single colored vertex. **(c)** Non-optimal weight 12 graph solution corrupting 12 of the data bits. **(d)** Optimal weight 3 graph solution corrupting 3 of the data bits. Note that after applying corrections corresponding to the blue edges, the data will be neither its original state nor its bit-inverse. This implies the described "correction" method is flawed.

E.1.3 Imperfect parity measurements

When parity measurements are imperfect, we can no longer process each round of parity information independently. For example, if the only error is in a parity measurement and we use the single-round algorithm of the previous section, any graph solution we choose will lead to disastrous corruption of the data, as shown in Fig. E.3.

We can fix our decoding algorithm by considering the effect of errors both in space and time. To do so we need to introduce the notion of a detection event, which corresponds to a change of a measurement parity in time. Given arbitrary values on the data bits and assuming no errors, each round of parity measurements will be the same as the previous round. When the parity changes, we know an error must have occurred nearby. For data errors, we see changes as described in the previous section. For parity measurement errors, there is first a change in the parity output, and then in the next cycle a change back to the original (correct) value. Thus a data error produces a pair of detection events in space (with single events at the boundaries),

while a parity-measurement error produces a pair of detection events in time. All these errors can be uniquely identified when sufficiently sparse.

We show an example of this behavior in Fig. E.4. A data error introduced at $t = 2$ gives a pair of detection events in space, while a measurement error at $t = 4$ gives a pair of detection events at times $t = 4, 5$. This figure also illustrates the basic idea of decoding the detection events into errors using the minimum-weight perfect matching algorithm. Here, a detection event is chosen and a region around it explored until another detection event or edge is found, whereby the two detection events are matched. The idea of the algorithm is that one should connect the red vertices to each other in pairs or to a graph boundary using the minimum total number of lines, where each line corresponds to the location of an error. Importantly, if errors are independent, patterns with fewer lines (errors) are more likely. The number of errors in a pattern is called its weight; an efficient algorithm solving this problem has existed since the mid-60's called minimum-weight perfect matching [**? ? ?**]. More details on the algorithm can be found in Ref.[**?**].

E.1.4 Time boundaries

When the last time boundary is encountered, the algorithm must have additional parity information to correctly match the detection events. As we must directly measure the bits anyway to check whether our decoding efforts were successful, we can use this data itself to compute the final round of parity measurements. As shown in Fig. E.5, this allows one to complete the graph, eliminating the future time boundary and enabling all detection events to be processed.

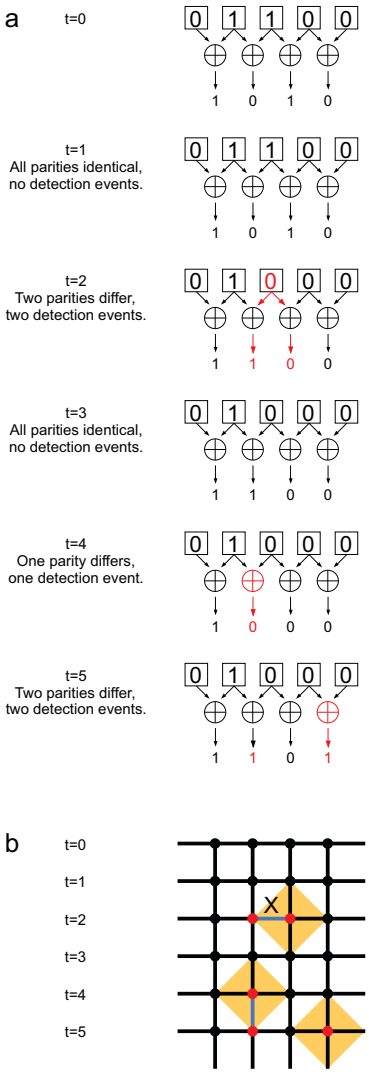


Figure E.4: **Processing imperfect parity measurements. (a)** Six rounds of imperfect parity measurements. Every time a parity differs from the previous round, a detection event is generated. **(b)** Corresponding graph problem. Red dots indicate detection events. Note that the graph now includes vertical lines to account for the possibility of parity measurement errors; a single parity measurement error manifests as a pair of sequential detection events. The matching algorithm works by selecting a detection event and exploring uniformly in all directions (orange diamonds) until a feature of interest is encountered. If the feature is another detection event, the two detection events are paired (blue lines). If the feature of interest is the future time boundary, we do not have sufficient information to correctly match the detection event, and need to wait for more data. Corrective data bit-flips are associated with horizontal matched edges. Vertical matched edges indicate parity measurement errors, and require no modification of the data.

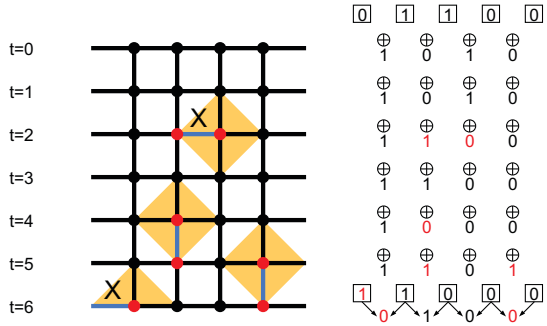


Figure E.5: **Complete processing of imperfect parity measurements.** At $t = 0$, the bits are initialized to 01100 and used to generate the parity measurements. No change at $t = 1$ indicates no initialization errors. At $t = 6$, to check whether storage has been successful, the bits are measured directly (but imperfectly). Any errors during measurement can be treated as errors occurring before a perfect round of measurements, implying no errors remain undetected even when using imperfect physical measurements. This data is used to generate the final parities. In this case, the error pair in the lower right corner is not matched to an edge (a data error), but is correctly identified as a measurement error.

Note that although the final data measurements are imperfect, we may model any such errors as coming from a data error in the previous round and treat the final measurements as error free.

A similar situation occurs for the beginning round, since there is nothing to compare to when computing the parity change at time step $t = 1$. We take the data as perfectly initialized to the desired value, so that errors in initialization are placed as data errors in the first round. The initial parities at $t = 0$ are then computed from this perfect initial state.

E.1.5 Two possible corrected outputs

By moving initialization and measurement errors to detection events, the initial and final states may be considered perfectly known. Logical errors arise only from decoding the detection

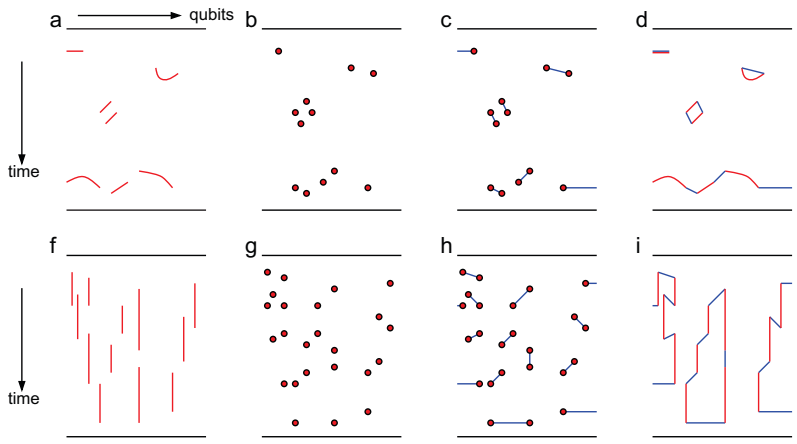


Figure E.6: **Topology of corrections.** **(a)** A large repetition code running for many cycles with error chains marked in red. **(b)** Corresponding detection events. **(c)** A minimum-weight matching of the detection events. **(d)** Errors and corrections plotted together. Successful corrections together with their associated errors form rings, or U-shapes starting and ending on a single boundary. Unsuccessful corrections together with their associated errors form chains connecting different boundaries. Given every horizontal edge corresponds to either an error or a correction, rings and U-shapes contribute a net identity operation. Chains connecting different boundaries contribute a net single bit-flip on every bit. Note that this is true even if the chain doubles back on itself, as then some bits will get an odd number of flips that will cancel down to a net single bit-flip. This is why, even in a physical system with realistic errors, after correction the output will either match the input or be its exact bit-inverse. **(e–i)** A more complex example focussing on extreme measurement error. In this case, errors and corrections cancel leading to successful storage.

events. As discussed in the case of only data errors, decoding gives the proper state or its logical error, the bit-inverse. This is also true for the general case of data and measurement errors, as illustrated graphically in Fig. E.6. It is interesting to note that the decoding process does not have to match exactly all error events, but only needs to correctly identify the totality of bit errors. As illustrated with the bottom line of Fig. E.6d, a logical error only occurs when a net error chain crosses the boundaries, which always produces a bit-inverse of the proper final state.

An advantage of this procedure is that the algorithm removes state preparation and mea-

surement errors (SPAM) to the same order as the error correction itself. This is a hallmark of fault-tolerance in that errors in every part of the quantum circuit are treated equally.

The idea that decoding gives the state or its bit-inverse is perhaps surprising, and although it is a mathematical statement, we have checked our decoding algorithm for consistency. For those who are concerned about using the final measurement data to help compute the errors in the final round, we note it is possible to compute errors without this data. In this case one would have greater errors because of imperfect matching of the final detection events, giving errors in both matching and the final measurement. As these additional errors come from the measurement, they are constant with changing the number of rounds, and thus the decay of the fidelity will be the same as for matching with the full measurement data.

E.2 Quantum repetition code: theory

Given qubits instead of bits, we need to be able to protect quantum superpositions from error. Qubits encode information in amplitude and phase, which can be expressed in terms of $\hat{X}$ and $\hat{Z}$ operators. Thus, errors can be expressed in terms of bit-flip ($\hat{X}$) and phase-flip ($\hat{Z}$) errors. However, detecting both types of errors simultaneously is nontrivial, as $[\hat{X}, \hat{Z}] \neq 0$. This can be overcome by constructing operators that measure the parity of two qubits. Take for example, the $\hat{X}_1\hat{X}_2$ and $\hat{Z}_1\hat{Z}_2$ operators:

$$\begin{aligned}
[\hat{X}_1\hat{X}_2, \hat{Z}_1\hat{Z}_2] &= (\hat{X}_1\hat{X}_2)(\hat{Z}_1\hat{Z}_2) - (\hat{Z}_1\hat{Z}_2)(\hat{X}_1\hat{X}_2) \\
&= \hat{X}_1\hat{Z}_1\hat{X}_2\hat{Z}_2 - \hat{Z}_1\hat{X}_1\hat{Z}_2\hat{X}_2 \\
&= (-\hat{Z}_1\hat{X}_1)(-\hat{Z}_2\hat{X}_2) - \hat{Z}_1\hat{X}_1\hat{Z}_2\hat{X}_2 \\
&= 0
\end{aligned} \tag{E.2}$$

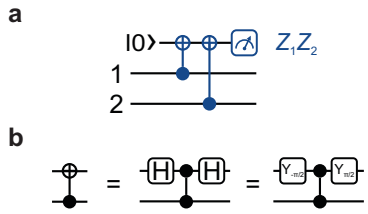


Figure E.7: **Parity measurement circuit.** (a) CNOT gates flip the state of the top ancilla qubit conditional on the state of the control qubits. The two CNOT gates and measurement will encode the eigenvalue of the $\hat{Z}_1\hat{Z}_2$ operator onto the state of the ancilla qubit, see text. (b) The CNOT gate can be decomposed into a controlled-phase (CZ) gate plus single-qubit gates. Decomposition can either consist of a Hadamard on the target qubit before and after the CZ, or a $Y_{-\pi/2}$ before and $Y_{\pi/2}$ after on the target qubit.

Thus, these multi-qubit operators can be used to detect the bit and phase parity of two qubits without knowing and collapsing the individual state of each qubit – analogous to the classical secret information example.

Consider the operator $\hat{Z}\hat{Z}$. This operator has the property that $\hat{Z}\hat{Z}|00\rangle = +|00\rangle$, $\hat{Z}\hat{Z}|01\rangle = -|01\rangle$, $\hat{Z}\hat{Z}|10\rangle = -|10\rangle$, and $\hat{Z}\hat{Z}|11\rangle = +|11\rangle$. This operator can detect changes in parity of the qubits; however, it cannot determine which qubit has flipped. To overcome this, we use a one-dimensional array of qubits and nearest neighbour parity operators, similar to the array of bits and XORs in the classical example. As we must protect qubits from both $\hat{X}$ and $\hat{Z}$ errors simultaneously, a fully protected state requires a two-dimensional array of qubits. Here, we design our experiment to focus on a chain of qubits and $\hat{X}$ errors only, as this is experimentally viable today.

We can construct the $\hat{Z}\hat{Z}$ operator through the use of quantum logic gates and measurement, as seen in Fig. E.7. Each controlled-NOT (CNOT) gate will flip the top ancilla qubit dependent on the state of the control qubit, just like the classical XOR operation. Thus, the state $|00\rangle$ will map the ancilla qubit to $|0\rangle$, the $|01\rangle$ or $|10\rangle$ states will map the ancilla to $|1\rangle$, and the $|11\rangle$

will have two flips that cancel, mapping the ancilla to $|0\rangle$. After the circuit is executed, the measured state of the ancilla qubit will encode the eigenvalue of the $\hat{Z}\hat{Z}$ operator.

E.3 Quantum repetition code: experiment

Our experimental implementation of the quantum repetition code consists of nine qubits total, five data qubits and four measure qubits. Limited memory in our control electronics restricts us to a maximum of eight parity measurement cycles. One complete run of our experiment consists of qubit initialization, between one and eight cycles of the physical gate sequence in Fig. E.26, data qubit measurement, then postprocessing to determine whether a logical error has occurred. This section focuses on the postprocessing, with emphasis on doing this with care to achieve the lowest possible logical error rate.

Any information pertaining to the physical performance of the device can be incorporated into the postprocessing to ensure that the very best possible corrections are suggested at the end of a run. We shall discuss five increasing levels of detail, with the final level of detail corresponding to the results reported in the main text. We do not claim to have exhausted available techniques.

E.3.1 Basic processing

There are a number of steps required to identify physical errors given the raw experimental output. Example data moved through each of these steps is shown in Fig. E.8. In the following sections, we will explain in detail how each of these steps is performed.

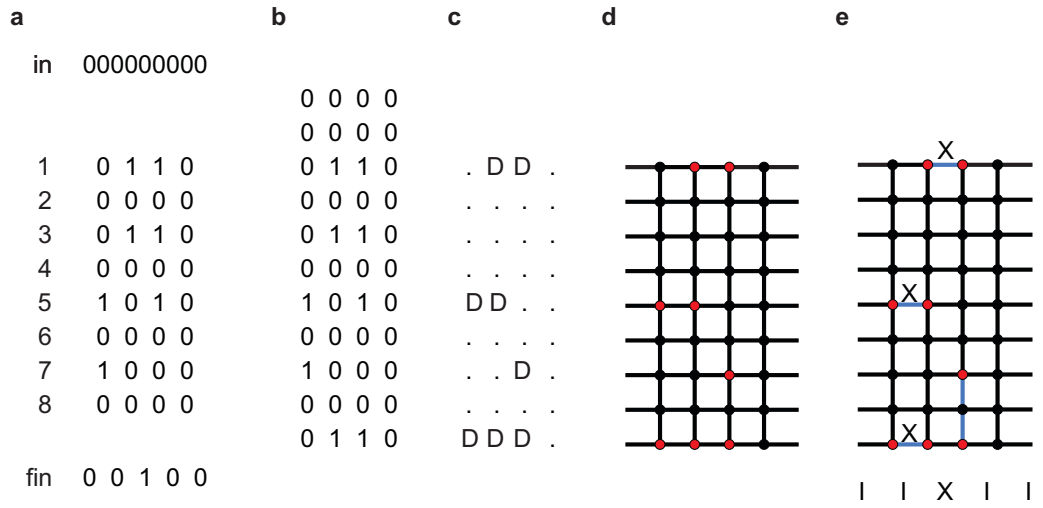


Figure E.8: **Flow of experimental data.** (a) Raw experimental data. (b) Raw experimental data appended by computed data to allow the first round of measurements to detect errors, and a final round of measure qubit outcomes inferred from the data qubit outcomes. (c) Data converted to detection events, indicated by D. (d) Detection events (red circles) placed onto graph which contains information about how errors propagate. (e) Detection events matched together (blue lines) to identify physical errors. The final two detection events on the third measure qubit, connected over three rounds, are an error chain. Final correction to recover data qubit input is shown at the bottom of the graph: $\hat{I} \otimes \hat{I} \otimes \hat{X} \otimes \hat{I} \otimes \hat{I}$.

(a) **Raw Data.** Before we can begin discussing more advanced processing, we must give a little more detail on how the experimental output is converted into detection events. Our experiment makes use of QND measurement and does not reinitialize measure qubits to $|0\rangle$ as was done in Fig. E.7a. This makes detection event identification more complex.

An example of experimental data gathered during an eight cycle run is shown in Fig. E.8a. Some postprocessing has already occurred, namely the conversion of the measurement microwave waveform into a best guess of the corresponding state. This postprocessing is described elsewhere (Section E.11). The first line "in" shows which state the nine qubits were intended to be initialized to. The five data qubits have been initialized to $|0_L\rangle$. The eight numbered lines show the output of each measurement qubit for each cycle, and the final line contains the experiment terminating data qubit measurements.

(b) **Additional simulated rounds.** In order for the first cycle of measurements to look for a change in parity, we generate two artificial rounds from a parity computation of the desired initial state. See Fig. E.8b. As the data qubits are initialized to either all 0's ($|0_L\rangle$) or all 1's ($|1_L\rangle$), the computed parity is all 0's. Likewise, an additional parity round is computed from the measurement of the final data.

(c) **Calculating detection events.** Consider the action of the parity measurement circuit in Fig. E.7a. If just one of the data qubits is in state $|1\rangle$, the value of the measure qubit will be flipped. If the data qubits have the same value, the measure qubit will be unchanged. In the absence of errors, QND measurement and no reinitialization therefore leads to two possible behaviors — alternating and constant, see main text. The presence of an

error is indicated by a change between alternating and constant behavior. For example, the sequence of measurement results 011 shows a change from alternating to constant behavior, and hence is associated with a detection event.

We desire a simple formula to identify detection events. Given a sequential pair of measurement results, m_{t-1}, m_t, we can use XOR to detect alternating $m_{t-1} \oplus m_t = 1$ and constant $m_{t-1} \oplus m_t = 0$ behavior. Given three sequential measurements, m_{t-2}, m_{t-1}, m_t, we can detect a change between alternating and constant behavior using $(m_{t-2} \oplus m_{t-1}) \oplus (m_{t-1} \oplus m_t) = m_{t-2} \oplus m_t$. In our experiment, a detection event is generated when a given parity measurement differs from the value reported two rounds ago.

The data converted into detection events is shown in Fig. E.8c. Our goal is to use the detection events to calculate bit-flips to apply to the final measurement results, with a success being recorded when the corrected output matches the input.

(d) **Detection events on error connectivity graph.** If we take no details of the underlying quantum circuit into account, and assume here for simplicity that we can only have parity measurement errors or data qubit errors between cycles of the repetition code, we can perform the classical postprocessing exactly as described in Subsection E.1.3 for the classical repetition code with imperfect parity measurements. This means using a graph with a square structure, with vertices at every potential detection event location. Vertices associated with actual detection events get colored red (Fig. E.8d). A potential data qubit error between a particular pair of repetition code cycles can be visualized as a horizontal

edge, a potential measurement error as a vertical edge.

(e) **Identifying errors.** As before, we use minimum-weight perfect matching to find a minimal set of errors reproducing the observed detection events. Blue edges show that, in this case, a unique minimum set of edges solves this problem. In general, the problem can have multiple solutions. Our processing deterministically but arbitrarily chooses one of these multiple solutions. In this instance, the correction suggested by the solution restores the observed output to the input, and the run has been successful. For the data set used for this paper, 3.35% of 9-qubit 8-cycle runs failed using this method of postprocessing.

For the remainder of this document, experimental data is organized as in Fig. E.8.

E.3.2 Data errors during the repetition code cycle

In the previous subsection, our processing assumed that data qubit errors could only occur between repetition code cycles. Despite the fact that this assumption is not, in fact, true, we gave a nontrivial example where the postprocessing succeeded. In reality, data qubit errors can occur at any time. Referring to Fig. 2 of the main text, an error on data qubit 2 during the third 20 ns window will be detected on measure qubit 3 immediately, but not on measure qubit 1 until the next cycle. Similarly, an error on data qubit 4 during the second 20 ns window will be detected on measure qubit 5 immediately, but not on measure qubit 3 until the next cycle. Finally, an error on data qubit 6 during the second or third 20 ns windows will be detected in the same cycle on measure qubit 5, but not on measure qubit 7 until the next cycle. Taking

into account these three new classes of error means including three new classes of edges in our graph problem.

Consider the following run.

```
in  101010101
      0 0 0 0
      0 0 0 0
1     0 0 0 0      . . . .
2     0 0 0 0      . . . .
3     0 0 0 0      . . . .
4     0 0 0 0      . . . .
5     0 0 0 0      . . . .
6     1 0 1 0      D . D .
7     0 1 0 1      . D . D
8     1 1 1 0      . D . .
fin 0 0 1 0 1      D D . .
```

Example E.1: An example run (real data) illustrating the benefit of considering data qubit errors during the repetition code cycle.

Using the basic postprocessing of Subsection E.3.1, the graph, solution, and suggested corrections are shown in Fig. E.9b, leading to failure. By including additional edges corresponding to data qubit errors during the repetition code cycle, a unique and better solution becomes available, resulting in a successful run. The 9-qubit 8-cycle logical error rate drops from 3.35% to

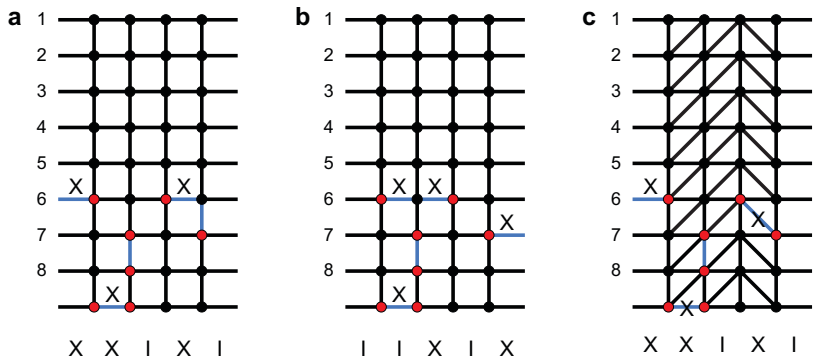


Figure E.9: Graph problems corresponding to Ex. E.1. Parts **(a)** and **(b)** show two possible solutions when assuming only basic errors as was done in Subsection E.3.1. Our code deterministically but arbitrarily chooses solution **(b)**, which for this experimental run was the wrong thing to do, resulting in failure. Without any additional information, solutions **(a)** and **(b)** both use 5 edges and are equally good. In part **(c)**, additional edges have been added to account for the possibility of data qubit errors during the repetition code cycle. When this is done, a unique 4 edge solution is found by postprocessing, successfully restoring the output to the input. Fewer edges means fewer errors and hence a more likely error pattern that matching should favor.

3.29% after including the additional edges. The improvement is small, as expected, as it is rare that errors occur in the precise windows required to generate behavior handled by these extra edges.

E.3.3 Weighted edges

In the previous Subsection, it was noted that there are few errors that can lead to a diagonally offset pair of detection events. It would seem reasonable, therefore, to make it less likely for postprocessing to choose diagonal edges when multiple options are available. Calculating the probability of each edge first requires an error model for every gate in the repetition code cycle. As a first pass, we choose to believe that every gate suffers errors well modeled by a depolarizing channel, and that gates of the same type suffer errors at the same rate.

Table E.1: Input error model.

Gate	Error
CZ	1%
X	0.1%
Idle (20 ns)	0.05%
Initialization	2.5%
Readout (measure qubit)	1.5%
Readout (data qubit)	3%.

We input the error rate associated with each operation in table E.1, determined from previous techniques [**?** **?**]. These operations are all that are required during our experiment. By studying exactly where and when every possible error is detected, we can determine the probability of every edge in our graph. We convert each edge probability p_i into a weight $w_i = -\ln p_i$ so that the weight of two consecutive edges is $w_i + w_j = -\ln p_i - \ln p_j = -\ln p_i p_j$. This ensures that minimum weight perfect matching [**?** **?** **?**] will consider two hypothetical detection events that can be matched either through independent single edges to a nearby boundary or to each other through two links to have two equally good matchings, as we wish, if the edge probabilities are equal.

Consider the following run.

```
in  101010101

      0 0 0 0

      0 0 0 0

1     0 0 1 0       . . D .

2     0 0 1 0       . . D .

3     0 0 0 0       . . D .
```

```
4     0 0 0 0       . . D .
5     0 1 0 0       . D . .
6     0 0 0 0       . . . .
7     0 1 0 0       . . . .
8     1 1 0 0       D D . .
fin 0 1 0 1 1       . D D .
```

Example E.2: An example run (real data) illustrating the benefit of carefully weighting each edge in the graph to reflect the actual probability of detection events being observed at the edge's vertices. Diagonal edges, in particular, are much less likely than horizontal and vertical edges.

Without taking the error rate of each gate into account, the corresponding graph problem has equally weighted edges and hence two equally acceptable solutions (Fig. E.10), only one of which leads to successful correction. With weights set according to their probability ($w = -\ln p$), low probability diagonal edges acquire high weights and the solution in Fig. E.10b, in this case the correct solution, becomes favored resulting in a successful run. The 9-qubit 8-cycle logical error rate drops from 3.29% to 2.897% through the use of weighted edges, a significant improvement. This is expected as a great deal of physical information has now been added to the graph, helping the postprocessor accurately distinguish between many previously degenerate solutions.

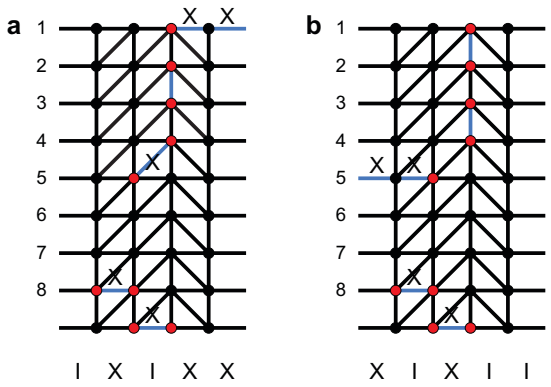


Figure E.10: Graph problem corresponding to Ex. E.2. With equally weighted edges, there are two equally good solutions with six edges. Our software deterministically chooses solution **(a)**, which in this case does not restore the output to the input, resulting in a failed run. By weighting edges according to their probability, diagonal edges get a much higher weight as they are low probability ($w = -\ln p$), penalizing their use. This breaks the degeneracy between these two solutions, favoring solution **(b)** and leading to a successful run.

E.3.4 Leakage

Superconducting qubits are not two-state systems, and leakage to other non-computational states can and does occur. When a measure qubit outputs 2, which will be the input to the next repetition code cycle, the next measurement result will be unreliable. We can only accurately predict what the next measurement result should be, even in the absence of errors, if the cycle begins with 0 or 1 on the measure qubit. This physical understanding can be fed into the postprocessing by setting the probability of a vertical edge to be 0.5 conditional on observing 2 at the end of the first cycle it is connected to.

Furthermore, a $|2\rangle$ on a measure qubit at the beginning of the repetition code cycle can induce errors on the neighboring data qubits. We observed that a neighboring data qubit in a $|1\rangle$ state is essentially randomized, whereas a neighboring data qubit $|0\rangle$ is mostly undisturbed. We can model this by setting the probability of the two horizontal edges associated with the

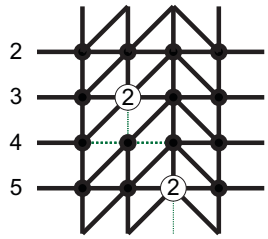


Figure E.11: When a measure qubit reports 2, the next measurement will be unreliable, and this knowledge can be accounted for by setting the probability of error for the next vertical edge to be 0.5, resulting in a very low weight link (narrow vertical line). Note that thinner lines are less weight, and therefore more likely. Neighboring data qubits, if they are in the $|1\rangle$, were observed to be strongly corrupted, which can be accounted for by setting the horizontal edges of the next cycle to have probability of error of 0.25 (slightly less narrow horizontal lines). These three high probability (low weight) edges will be strongly preferred by the matching algorithm, more reliably pairing up cascades of detection events resulting from leakage.

next repetition code cycle to 0.25. These two reweighting rules are depicted in Fig. E.11

Consider the following run.

```
in  101010101

    0 0 0 0

    0 0 0 0

1   0 0 0 0       . . . .

2   0 0 0 0       . . . .

3   0 2 1 0       . D D .

4   1 1 1 0       D D D .

5   0 1 2 0       . . . .

6   1 0 2 0       . D . .

7   0 2 2 2       . . . D

8   1 1 2 2       . D . D
```

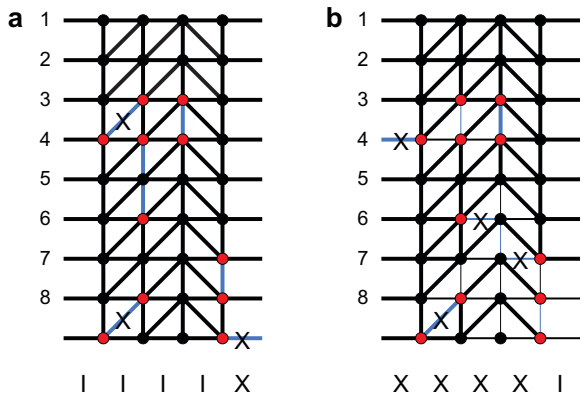


Figure E.12: **(a)** matching of Ex. E.3 without accounting for leakage. **(b)** leakage is well handled by lowering the weight of nearby edges following detected leakage. These low weight edges encourage detection events generated as a result of leakage to be paired together leading to a higher probability of successful correction. In this example, after reweighting the suggested corrections successfully return the output to the input.

```
fin 0 0 0 0 1     D . . D
```

Example E.3: An example run (real data) illustrating the benefit of lowering the weights of future edges after a measure qubit reports 2. Note that a 2 is treated as a 1, plus reweighting.

A large region of the data has been corrupted by leakage. Without accounting for leakage, the matching shown in Fig. E.12a is found, and the corresponding corrections are unsuccessful. With edge reweighting, detection events arising from leakage are preferentially matched to each other, in this instance leading to successful correction. Leakage plays a significant role in superconducting circuits. Including $|2\rangle$ state intelligence in the measurement and postprocessing reduces the 9-qubit 8-cycle logical error rate from 2.897% to 2.414%, a significant improvement.

E.3.5 Fine-tuning the error model

In Subsection E.3.3, we chose simple depolarizing error rates for each type of gate and assumed that every gate of the same type had the same error rate. We shall now relax that assumption, assuming only that every gate of the same type applied to the same qubits will have the same error rate.

To begin, we associate a variable with each geometrically- or weight-distinct edge. The set of variables is represented graphically in Fig. E.13. Geometrically-distinct means connecting different sets of qubits, or being horizontal versus diagonal. Weight-distinct means that while the edges may be geometrically-identical, they have different weight. The final row of horizontal edges has a different weight to every other row as the final 3% error data qubit measurements contribute to the generation of detection events associated with this row only, lowering their weights below that of other horizontal lines. Similarly the 2.5% initialization error contributes to the first row of horizontal and vertical edges only. The remaining green edges control the rest of the graph, with a small number of variables controlling the repetitive weight structure.

Our primary data file contains 90,000 runs for each possible number of cycles from 1 to 8 and both $|0_L\rangle$ and $|1_L\rangle$ as input, for a total of 1.44 million runs. When our postprocessing is run on this file using the initial depolarizing error model, 18,750 runs lead to logical errors. We can use the total number of logical errors as a fitness metric and optimize the edge variables defined above for a total logical error count. We make use of the standard Nelder-Mead algorithm to perform this optimization. Optimizing the error model in this manner reduces the 9-qubit 8-cycle logical error rate from 2.414% to 2.246%.

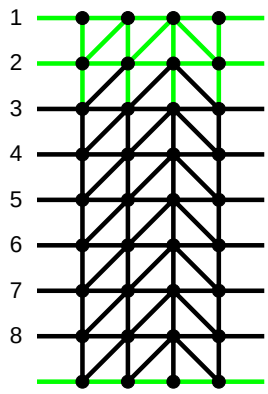


Figure E.13: Green edges denote geometrically- or weight-distinct edges. Every black edge is equivalent to a green edge by time translation. The relatively low number of unique edges is due to the cyclic nature of the repetition code, and hence the cyclic nature of error propagation. Two edges for coping with leakage are also optimized, but not shown as they are dynamically added when leakage is detected.

Optimizing using the entire data set is cheating, however, since it doesn't guarantee good performance on unseen data. To combat this, we first split the full dataset in two parts. The chosen split was between even and odd entries in the dataset, which ensures full sampling over the time of the experiment and input states. Second, we use these two sets to train two error models. We obtain logical error rates of 2.222% for the even, and 2.254% for the odd dataset. Note that the logical error rates differ only in the third significant figure, illustrating that the optimized performance is not strongly dependent on the dataset.

Third, to further test the validity of these error models, we test their performance on the datasets they have not seen. We obtain logical error rates of 2.300% for the even error model on the odd dataset and 2.243% vice versa. We note that the performance here is very close to the full dataset, and that indeed, the error model trained on odd data, performs even slightly better when used on the even dataset. These results underline that the error models are nearly

Table E.2: Summary of logical error rates.

algorithm	logical error rate
basic errors only	3.350 %
data errors during the cycle	3.290 %
weighted edges	2.897 %
leakage	2.414 %
fine-tuning the error model	2.300 %

equivalent, and therefore can be used to infer a reliable logical error rate. We took the worst results, namely optimizing on even and running on odd data, to be the foundation of the discussion in Chapter 8.

In summary, our most basic postprocessing led to a logical error rate of 3.35%, and our final hardware-optimized postprocessing reduced this to 2.3%, a 30% reduction. Our stepwise optimization are summarised in table E.2. In the next section, we shall see that there is a better figure of merit characterizing the performance of a fault-tolerant quantum error correction system. We think that this is a significant test of the theory that better error models give a lower logical error rate.

E.4 Error suppression factor Λ - figure of merit for fault-tolerant quantum error correction

This Section discusses a universal figure of merit characterizing the performance of any fault-tolerant topological quantum error correction system, meaning the totality of both the quantum hardware and the classical postprocessing. Clearly, the 9-qubit 8-cycle logical error rate we discussed in the previous Section is highly specific to our current experiment, and unlikely to

be appropriate for comparison with other hardware or experiments. We seek instead a generic figure of merit characterizing how rapidly errors are suppressed as qubits are added to the system. This means studying and comparing the performance of different orders of fault-tolerance.

In the main text we defined n-th order fault-tolerance to mean a system guaranteeing correction of any combination of n errors. Our full 9-qubit experiment is 2nd-order fault-tolerant to X errors. There are three possible 5-qubit subsets that are 1st-order fault-tolerant. In an effort to make the most reliable extrapolations to higher orders, instead of separately running the three possible 5-qubit subsets, we infer the performance of these subsets directly from the full 9-qubit data. This simply means discarding appropriate columns from a larger 9-qubit run. We believe this will give a more accurate estimation of Λ than running three separate 5-qubit experiments, which is experimentally taxing.

For example, the performance of the first 5 qubits can be inferred from 9-qubit data as follows.

in	000000000	00000
1	0 1 1 0	0 1
2	0 0 0 0	0 0
3	0 1 1 0	0 1
4	0 0 0 0	0 0
5	1 0 2 0	1 0
6	0 0 0 0	0 0

```
 7    1 0 0 0          1 0

 8    0 0 0 0          0 0

fin 0 0 1 0 0        0 0 1
```

Example E.4: Left, full 9-qubit run data. Right, the same data restricted to the first five qubits.

This works because the repetition code is topological – its structure is transversely invariant as it grows and its classical processing is local on average. The extra piece of unused repetition code in the above example fundamentally just interacts with the rightmost data qubit introducing a small amount of additional error on this qubit. This in no way conceptually changes the performance or required postprocessing of the leftmost five qubits.

Ideally, a well-constructed topological quantum error correction system should have a logical error rate that scales as $\epsilon \sim 1/\Lambda^{n+1}$. The universal figure of merit Λ specifies how much the logical error rate will go down if the order n of fault-tolerance is increased. For Λ to be well-defined, the system must operate with sufficiently low error rates to become more reliable as it grows. Note that it has been traditional to focus on a threshold error rate in the literature, as this is a theoretical quantity that is straightforward to calculate through simulation. The threshold error rate is not, however, a terribly meaningful experimental quantity, as its measurement would require variable error rate gates, necessitating the artificial insertion of noise. Rather, Λ measures how far below threshold error rate a system is, without requiring an explicit measurement of this threshold value.

To obtain an estimate of Λ, one requires, at a minimum, a single system capable of demonstrating 1st- and 2nd-order fault-tolerance. It is not sufficient to compare a single qubit to the

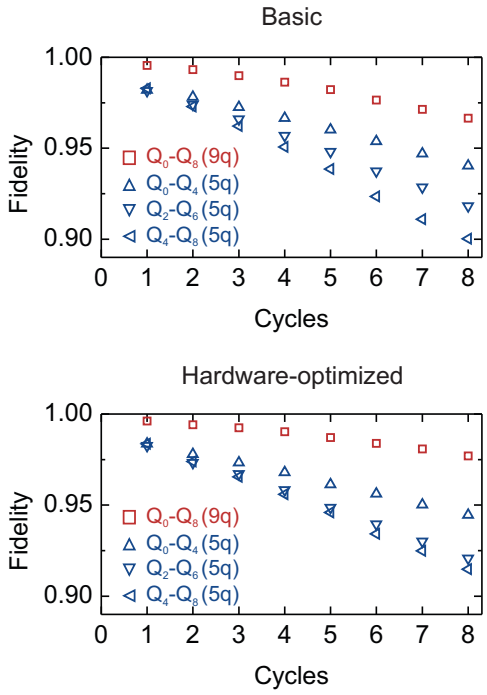


Figure E.14: **Comparison of basic data processing and hardware-optimized data processing.** (top) Fidelity of basic data processing for each combination of five qubits, and nine qubits $vs.$ cycle number. (bottom) Fidelity of hardware-optimized data processing for each combination of five qubits, and nine qubits $vs.$ cycle number.

performance of 1st-order fault-tolerance as a bare qubit may have a very long memory time, but high-error gates, leading potentially to poorer performance of the error corrected system than a bare qubit, despite the system being below threshold ($\Lambda > 1$).

In Fig. E.14 we plot the three possible five qubit logical error rates and the nine qubit logical error rate for both basic and hardware-optimized postprocessing. To calculate Λ, we first calculate the average five qubit logical error rate and take the ratio of average five qubit error rate to nine qubit error rate. Given the accumulation of leakage, as discussed in the main text, we find an increase in error rate with cycle number, thus degrading Λ with cycle number (Fig. E.15). A truly scalable system would asymptote to a constant value of $\Lambda > 1$. Our data is

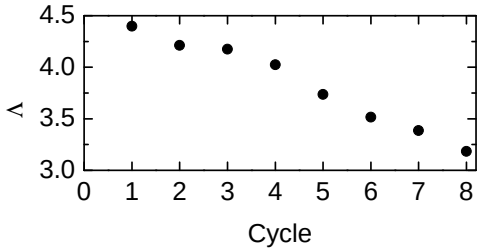


Figure E.15: **Changing Λ with cycle number.** We find that Λ, the ratio of five qubit to nine qubit error rate, decreases with increasing cycle number, see section E.7.

currently insufficient to know if this will occur.

E.5 Physical leakage mechanisms

From previous experiments, we measure a CZ gate to have $1 - 2 \cdot 10^{-3}$ population leakage, and a single qubit gates to have at most $\sim 10^{-4}$, determined in Ref. [**?**]. Empirically, we find that dispersive readout at higher powers can cause state transitions from $|1\rangle$ to $|2\rangle$, and we suspect this is the primary source of leakage.

E.6 Decomposing the failure rate

The state of the data qubits and the decomposition of the error rate is displayed in Fig. E.16 for eight cycles of the nine qubit repetition code, with 90,000 statistics. The data qubit input state was $|00000\rangle$, hence any deviation of this state amounts to a bit error. In Fig. E.16a, we find that the number of bit errors in the final data qubit state varies from a ratio of 0.33 for no error to a ratio of $6.9 \cdot 10^{-4}$ for five bit-flips. The ratio follows an exponentially decreasing

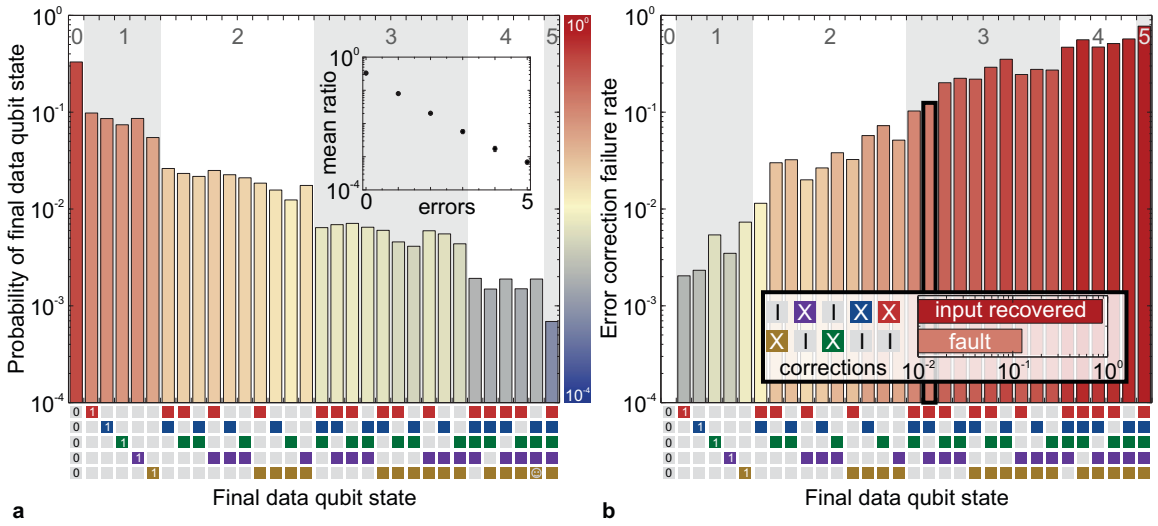


Figure E.16: **Decomposition of probability and failure rate for the nine qubit repetition code after eight cycles for input state $|00000\rangle$.** (a) Histogram of the final data qubit state, ordered with the number of bit-flips. Coloured squares indicate $|1\rangle$ states. Inset: averaged ratio versus number of bit-flips errors. (b) Error correction failure rate versus final data qubit state. The inset shows an example for the final data qubit state $|01011\rangle$ (highlighted): in 87.5% of the cases the minimum weight perfect matching gives the operator $\hat{I} \otimes \hat{X} \otimes \hat{I} \otimes \hat{X} \otimes \hat{X}$ which changes the final data qubit state into $|00000\rangle$, successfully recovering the input state. In 12.5% of the cases however, the inverse operator is given, leading to failure. We find no failure for the final data qubit state $|00000\rangle$. The experiment was repeated 90000 times.

trend with number of errors, indicating correlated bit-flips are nearly absent. We note that when running the repetition code for a large number of cycles, the physical state of the data qubits randomizes, and asymptotically approaches a uniform distribution.

At the end of each run, the minimum weight perfect matching algorithm gives one of two outcomes: the operator to flip the data qubits back to the input state (success), or to the bitwise inverse (failure). The error correction failure rate is plotted in Fig. E.16b as a function of final data qubit state. We find no failure for the final data qubit state $|00000\rangle$. We find a failure rate ranging from $2 \cdot 10^{-3}$ to $11 \cdot 10^{-3}$ for the case of one bit-flip in the final data qubit state, and a general trend of increasing failure rate with increasing number of data qubit bit-flips.

The case for final state $|01011\rangle$ is highlighted, showing that in most cases the matching algorithm gives the correct operator $\hat{I} \otimes \hat{X} \otimes \hat{I} \otimes \hat{X} \otimes \hat{X}$, which changes the final state back to the input state. In 12.5% of these cases the matching algorithm gives the inverse operator, leading to failure. The final data qubit state may obscure measurement errors as well as multiple bit-flips which cancel each other.

E.7 Increasing number of detection events with cycle

We observe an increasing pattern of detection events with cycle number. We attribute this to two main causes: state leakage, and measure qubit energy relaxation.

E.7.1 State leakage

State leakage, the population of the non-computational $|2\rangle$-state, is shown to grow with repetition code cycle, see Fig. E.17a. This population reduces the fidelity of the CZ entangling gate, leading to an increase in detection events.

In Fig. E.17b, we plot the fraction of detection events, defined by the number of detection events divided by the total number of possible detection events, averaged over both input states and all eight-cycled runs. We plot this fraction when using standard $|0\rangle$ and $|1\rangle$ state discrimination (black), as well as when post-selecting out the detection events for $|2\rangle$ leakage (red).

For runs without $|2\rangle$-events, we notice I) a significantly reduced amount of detection events, II) nearly no dependence of detection events for measure qubits Q_1 and Q_3 on cycle number, III) a reduced dependence (slope) for qubit Q_3. The data suggest that non-computational leakage is a significant contribution to the amount of detection events and the increase with cycle number.

E.7.2 Energy relaxation

We attribute the remainder of the increase in detection events to measure qubit energy relaxation, indirectly induced by randomization of data qubits. For input states $|0_L\rangle$ and $|1_L\rangle$, the measure qubits end in the state $|0\rangle$ at each cycle (see Fig. 1 of the main Letter). For our system $|0\rangle$ is highly robust. However, after several cycles some of the data qubits are likely flipped, and some of the measure qubits will switch between $|0\rangle$ and $|1\rangle$ with every round; $|1\rangle$ is susceptible

to energy relaxation.

A clear indication of this error mechanism is shown in Fig. E.17c, where we plot the fraction of detection events as a function of cycle number. The fraction of detection events alternate with every round. This is compatible with energy relaxation affecting the measure qubits, which switch between $|0\rangle$ and $|1\rangle$ with every round. The rise is then compatible with the increasing randomization of data qubits with cycle number. For either input state we find quantitatively a very similar behaviour, further supporting this interpretation.

E.8 Sample fabrication

Devices are fabricated identically to Ref. [**?**]. The fabrication details are reproduced here for convenience.

The devices are made in a process similar to the fabrication steps outlined in Ref. [**?**], with an important improvement: we have added crossovers to suppress stray microwave chip modes by tying the ground planes together with low impedance connections. Otherwise, the many control wires in our chip could lead to segmentation of the ground plane, and the appearance of parasitic slotline modes [**?**]. The device is made in a five-step deposition process, (I) Al deposition and control wiring etch, (II) crossover dielectric deposition, (III) crossover Al deposition, (IV) Qubit capacitor and resonator etch, (V) Josephson junction deposition. The qubit capacitor, ground plane, readout resonators, and control wiring are made using molecular beam epitaxy (MBE)-grown Al on sapphire [**?**]. The control wiring is patterned using lithography and etching with a BCl_3/Cl_2 reactive ion etch. A 200 nm thick layer of SiO_2 for

the crossover dielectric is deposited in an e-beam evaporator, followed by lift-off. We fabricate crossovers on all the control wiring, using a SiO_2 dielectric that has a non-negligible loss tangent. An in-situ Ar ion mill is used to remove the native AlO_x insulator, after which a 200 nm Al layer for the crossover wiring is deposited in an e-beam evaporator, followed by lift-off. We used 0.9 μm i-line photoresist, lift-off is done in N-methyl-2-pyrrolidone at 80°C. A second

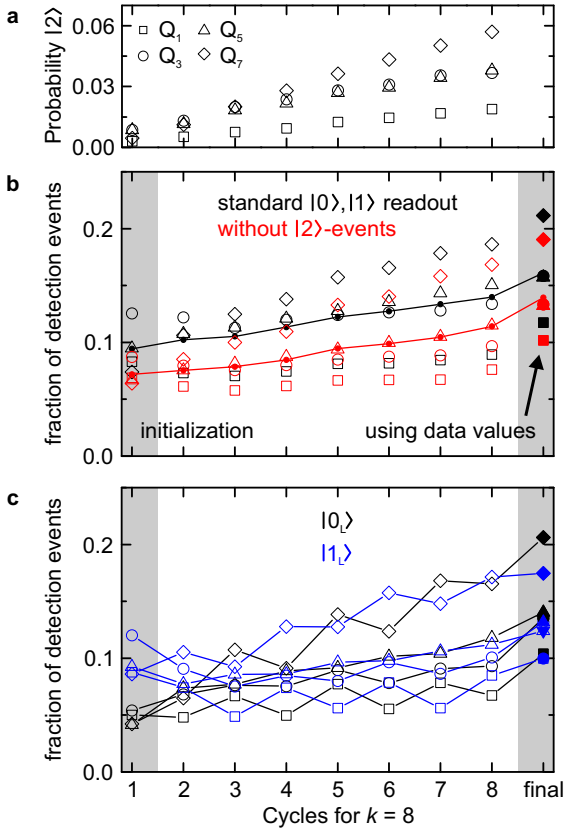


Figure E.17: **State leakage and fraction of detection events for $k = 8$.** (a) State leakage, $|2\rangle$-state population, for the four measurement qubits. Note that increase in measured $|2\rangle$ state population can come from misidentification of increasing $|1\rangle$ state population. (b) Fraction of detection events for the four measure qubits as a function of cycle number for standard $|0\rangle$ and $|1\rangle$ state discrimination (black), and without $|2\rangle$-events. Solid lines denote the averages. (c) Fraction of detection events for data qubit input states $|00000\rangle$ and $|11111\rangle$, showing a clear alternating pattern of increased detection events when measure qubits are likely in the $|1\rangle$ state. Used data is without $|2\rangle$-events.

BCl_3/Cl_2 etch is used to define the qubit capacitor and resonators; this step is separate from the wiring etch to prevent the sensitive capacitor from seeing extra processing. Lastly, we use electron beam lithography, an in-situ Ar ion mill, and double angle shadow evaporation to deposit the Josephson junctions, in a final lift-off process. See Ref. [**?**] for details.

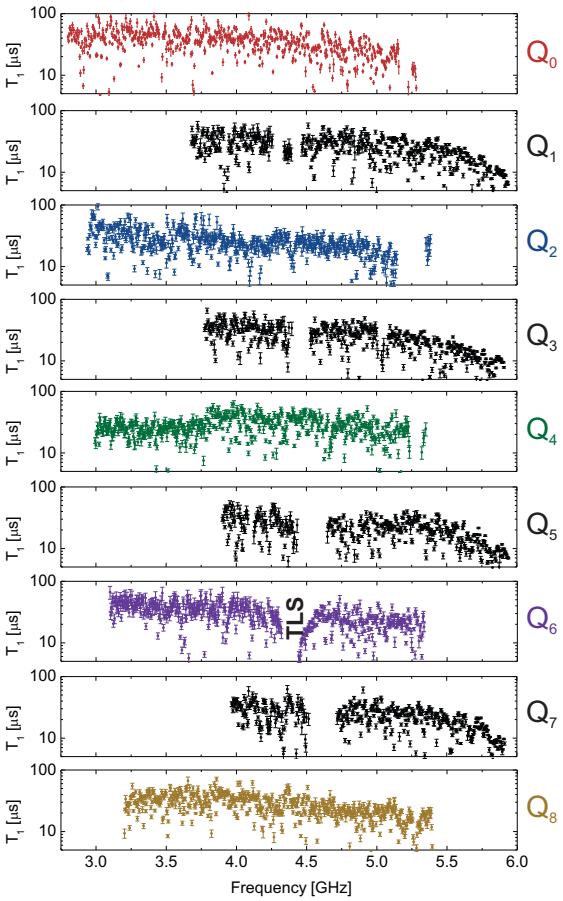


Figure E.18: **Energy Relaxation.** T_1 for all nine qubits as a function of frequency. Xmon qubits have a 24 μm gap and 24 μm center width for the capacitor. Missing data due to coherent swapping with other qubits or low quality fits. Two-level state (TLS) seen in the spectrum of Q_6.

E.9 Qubit coherence

We measure T_1 values for all nine qubits as a function of frequency in Fig. E.18. We consistently find values in the 20-50 μs range.

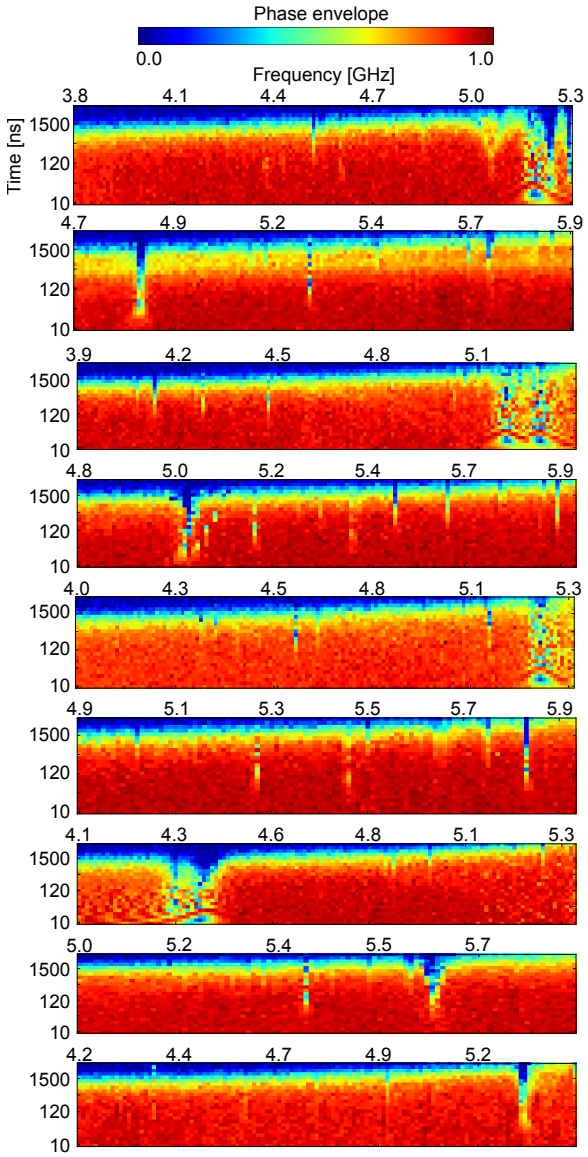


Figure E.19: **Dephasing.** Ramsey experiment for all nine qubits as a function of frequency. Phase envelope is the length of the Bloch vector on the XY plane. Oscillating features are interactions with coupled qubits.

In Fig. E.19 we perform Ramsey experiments as a function of frequency. We find Ramsey $1/e$ times of 15 μs near the flux-insensitive point, and Ramsey $1/e$ times varying between 2 and 5 μs at frequencies away from the flux insensitive point.

E.10 Measurement: Readout and bandpass filter design

The readout circuitry of the device is highlighted in Fig. E.20. Qubits (blue) are coupled capacitively to individual readout resonators (purple). The readout resonators are coupled inductively to a bandpass filter (green), which is weakly coupled to the input (white, coupling quality fac-

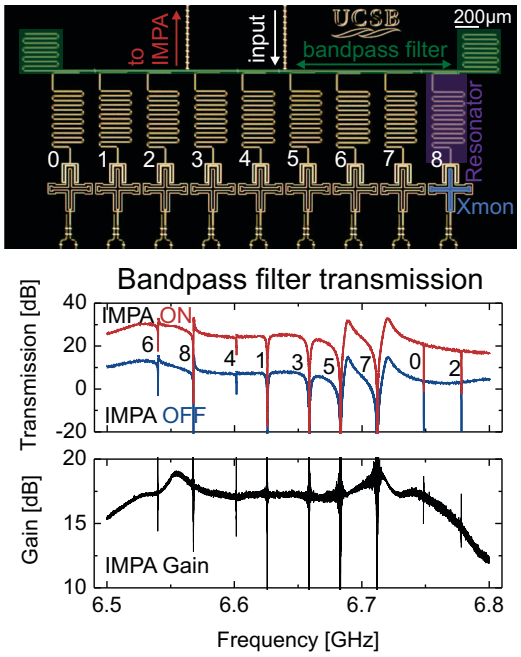


Figure E.20: **Bandpass filter and readout design.** Nine qubits (blue), with individual readout resonators (purple) are coupled to single half wave ($\lambda/2$) bandpass filter (green). The filter isolates qubits from the 50 Ω environment. A high bandwidth, high saturation power impedance matched parametric amplifier (IMPA) allows for frequency domain multiplexed readout using a single readout line.

tor $Q = 300$) and strongly coupled to the output (red, coupling quality factor $Q = 33$). At the output an impedance matched parametric amplifier (IMPA) acts as first stage amplifier [**?**]. The high bandwidth and saturation power is critical for system performance.

The bandpass filter isolates qubits from the 50 Ω environment formed by the readout lines. The bandpass filter (BPF) design used in this device is similar to Ref [**?**]. Here, the BPF is designed for a bandwidth of roughly 220 MHz, so placing nine readout resonators in this band with 30 MHz spacing requires a high level of precision. We use a design which geometrically enforces minimal frequency difference between resonators and filter:

- The filter used here is a half wave (λ/2) resonator. This provides more space for coupling all qubits to the same filter.

- The input and output lines are coupled using voltage taps, which do not shift the filter frequency. A parallel plate capacitor, for example, could cause the frequency to shift if the dielectric thickness is unreliable.

- The coplanar waveguide geometry of the filter is chosen identical to that of the readout resonators. Therefore the kinetic inductance changes the frequencies of filter and resonators equally.

- The coupling capacitor between readout resonator and qubit has identical geometry for both data and measure qubits. This ensures a proper frequency spacing of readout resonators, as the self-capacitance, and therefore the pull on the readout resonator frequency, is identical.

- We chose a small resonator-filter coupling for data qubits to reduce measurement induced dephasing.

- We time-multiplex the readout of data and measure qubits to achieve high-fidelity readout for all nine qubits. The IMPA provides a gain of 15-18 dB, with saturation power around -100-110 dBm for the entire band. While the saturation power is high, it is a limitation for simultaneous readout of all nine qubits, necessitating time-multiplexing: We read out all five data qubits simultaneously, and all four measure qubits simultaneously.

E.11 Multi-state readout

The demodulated inphase (I) and quadrature (Q) data from individual readout shots for preparing the $|0\rangle$, $|1\rangle$, and $|2\rangle$ states are shown in Fig E.21 for each measure qubit. After accumulating statistics, the location in IQ space for the ideal $|0\rangle$, $|1\rangle$, and $|2\rangle$ states are determined. State discrimination is performed by calculating which ideal state is closest in IQ space to a measured value.

E.12 Device parameters

The device parameters are listed in table E.3. Note that the coupling rate g is defined such that strength of the level splitting on resonance (swap rate) is $2g$ (Ref. [**?**]).

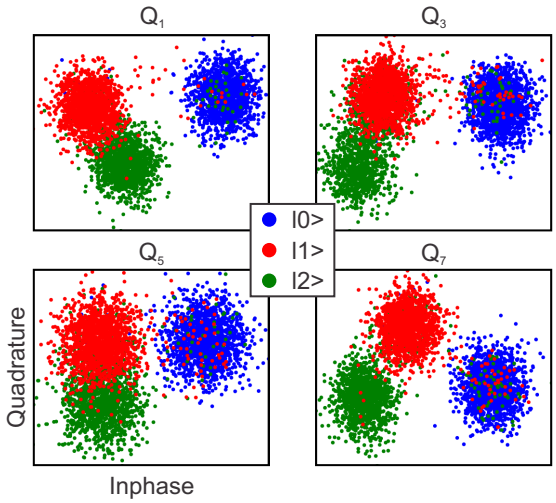


Figure E.21: **Three state readout for measure qubits.** Individual measurement shots for preparing the $|0\rangle$, $|1\rangle$, and $|2\rangle$ states for each measure qubit.

Table E.3: Parameters for the device when running the repetition code. f are frequencies. η is qubit nonlinearity. g is coupling strength. κ is resonator leakage rate.

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
Qubit frequencies and coupling strengths									
f_{10}^{max} (GHz)	5.30	5.93	5.39	5.90	5.36	5.94	5.33	5.91	5.39
$\eta/2\pi$ (GHz)	-0.230	-0.216	-0.229	-0.214	-0.227	-0.214	-0.242	-0.212	-0.225
f_{10}^{idle} (GHz)	4.3	5.18	4.43	5.28	4.49	5.40	4.60	5.46	4.7
f_{res} (GHz)	6.748	6.626	6.778	6.658	6.601	6.687	6.540	6.718	6.567
$g_{\text{res}}/2\pi$ (GHz)	0.110	0.128	0.111	0.109	0.110	0.110	0.098	0.111	0.104
$g_{\text{qubit}}/2\pi$ (MHz)	13.8		14.1		15.4		14.4		
$g_{\text{qubit}}/2\pi$ (MHz)		14.5		14.4		14.6		15.6	
$1/\kappa_{\text{res}}$ (ns)	675	69	555	30	1144	36	590	28	473
Readout (RO) parameters									
RO error	0.015	0.004	0.067	0.007	0.048	0.013	0.017	0.011	0.018
simult. RO error		0.004		0.012		0.022		0.013	
separation error		$4 \cdot 10^{-6}$		$2 \cdot 10^{-5}$		$2 \cdot 10^{-3}$		$2 \cdot 10^{-3}$	
Thermal $\|1\rangle$ pop.	0.013	0.007	0.028	0.01	0.037	0.018	0.012	0.009	0.012
RO pulse length (ns)	800	160	800	300	800	300	800	300	800
RO demodulation length (ns)	800	560	800	460	800	460	800	460	800
$f_{10,\text{RO}}$ (GHz)		5.46		5.31		5.40		5.54	
resonator n_{photons}		37		18		10		14	
Gate parameters									
Single qubit gate error		0.0006		0.0009		0.001		0.001	
X_π length (ns)	25	20	25	20	25	20	25	20	25
CZ length (ns)	45		45		45		45		
CZ length (ns)		45		45		45		45	
Qubit lifetime at idling point									
T_1 (μs)	26.3	24.7	39.2	21.3	41.1	19.1	22.0	28.1	18.6

E.13 Preservation of two-qubit gate fidelity when scaling up

We quantify the fidelity of multi-qubit operating by using two-qubit randomized benchmarking (RB), shown in Fig. E.22. See Ref. [**?**] for details on the implementation of Clifford-based randomized benchmarking. The reference decay from performing two-qubit Cliffords C_2 is a metric for system performance, as it contains simultaneous single and two-qubit gates (each C_2 contains on average $\frac{33}{4}$ single qubit gates and $\frac{3}{2}$ CZ gates). We have performed two-qubit randomized benchmarking on qubits Q_4 and Q_5 as a characteristic pair. We find an average error per two qubit Clifford C_2 of 0.0191, which is close to the result of 0.0189 obtained for the five qubit chip in Ref. [**?**]. This shows that gate performance was maintained while scaling up to nine qubits and integrating high-fidelity measurement.

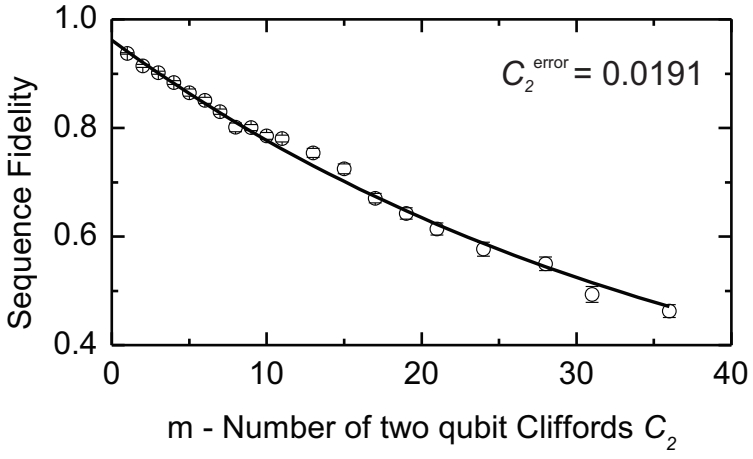


Figure E.22: **Two-qubit randomized benchmarking data.** The reference curve for two-qubit Cliffords C_2 for Q_4 and Q_5. The observed error rate is consistent with a simultaneous single qubit gate fidelity > 0.999 and a CZ gate fidelity > 0.99.

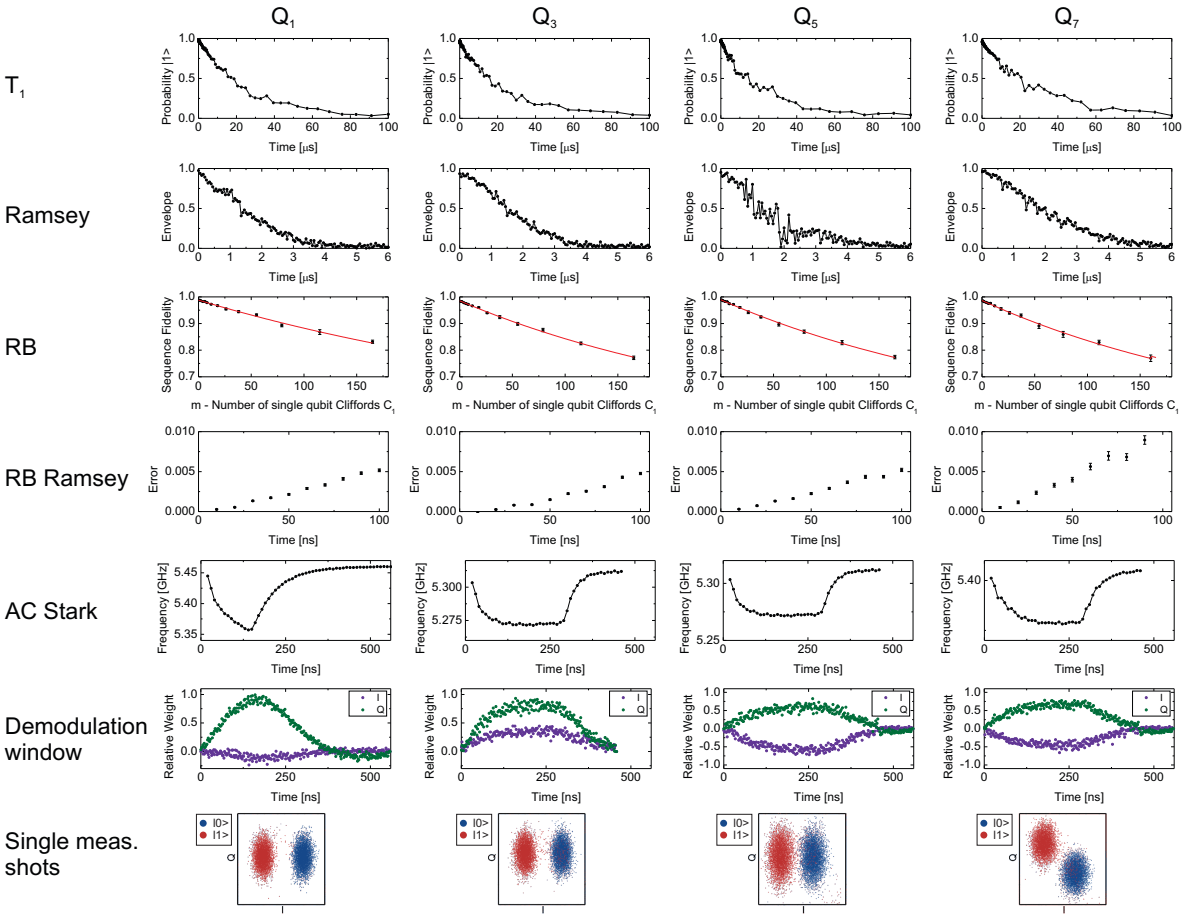


Figure E.23: **Measure qubit characterization.** We measure coherence, gate and readout properties on the four measure qubits. We find T_1 values in the 20-30 μs range for these operating points. Ramsey 1/e decay times are in the 2-3 μs range. Single qubit gate fidelities all meet or exceed 0.999. The RB Ramsey [**?**] experiment shows that the Ramsey 1/e time is not a limitation to gate fidelity. The AC stark shift and optimal demodulation windows displayed are used to achieve separation fidelities > 0.998.

E.14 Measure qubits in detail

We carefully characterize the four measure qubits in this device in Fig. E.23. These qubits simultaneously combine long coherence with high fidelity gates and readout. For a full listing of parameters, see table E.3.

E.15 Z control: crosstalk

We measure a crosstalk between the frequency Z control lines and qubits that is small, approximately $0.1 - 0.9\%$. The crosstalk matrix M_Φ is shown below, defined as: $\Phi_{\text{actual}} = (1 + dM_\Phi)\Phi_{\text{ideal}}$, with Φ the flux threaded through each qubit's superconducting quantum interference device (SQUID) loop. The magnitude of the crosstalk drops off with distance between lines and the sign depends on the routing direction of the wires, as expected.

$$dM_\Phi \approx 10^{-3} \begin{pmatrix} 0 & 2 & -1 & -2 & -3 & -1 & -2 & -2 & -2 \\ 2 & 0 & -5 & -2 & -3 & -1 & -2 & -1 & -2 \\ 2 & 5 & 0 & -5 & -3 & -2 & -3 & -1 & -2 \\ 1 & 4 & 2 & 0 & -7 & -3 & -3 & -2 & -2 \\ 1 & 4 & 2 & 2 & 0 & -8 & -5 & -2 & -2 \\ 1 & 3 & 2 & 2 & 0 & 0 & -9 & -4 & -2 \\ 1 & 3 & 1 & 2 & 0 & -1 & 0 & -7 & -2 \\ 1 & 2 & 1 & 2 & 0 & -1 & -5 & 0 & -6 \\ 1 & 2 & 1 & 2 & 1 & -1 & -7 & -1 & 0 \end{pmatrix}$$

E.16 Running the repetition code

E.16.1 Controlled-phase gates with nearest neighbor coupling

Our system consists of a linear array of qubits with nearest neighbor coupling g, where multi-qubit Ω_{ZZ} interactions are turned on and off by frequency tuning [**?**]. As each qubit is coupled to more than one neighbor, operation frequencies must be carefully considered to prevent unwanted interactions.

There are three effects that must be simultaneously considered, see Fig. E.24: I) The interaction which enables the CZ gate, by bringing the $|11\rangle$ and $|02\rangle$ close to resonance, needs to be turned off by detuning (Δ) the qubits; this interaction turns of quadratically for $\Delta \gg g$. II)

During a CZ gate, other neighboring qubits must be detuned to mitigate stray interactions, see Fig E.24b. III) Next-nearest neighboring qubits have a small parasitic coupling (on the order of $g_{i,i+2}/2\pi = 0.75$ MHz). Therefore, while idling they effectively perform a parasitic CZ. To minimize this interaction we detune next-nearest neighbor qubits by $\eta/2 \approx 100$ MHz which is sufficient to prevent coupling of $|10\rangle$ to $|01\rangle$ and $|11\rangle$ to $|02\rangle$, see Fig. E.24c.

With these three effects in mind, we consider how to perform the CZ portion of the repetition code. Each repetition code cycle consists of a CZ between each pair of neighboring qubits; 8 CZ gates for a 9 qubit array. This sequence can be executed, in principle, in the time of two CZ gates by performing 4 CZ gates in parallel at at time, this is technically demanding and

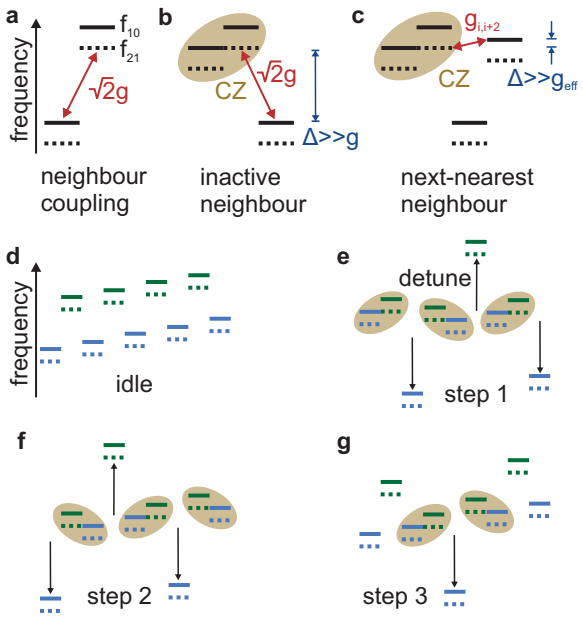


Figure E.24: **CZ interactions.** (a) Neighboring qubits are coupled with $\sqrt{2}g$ coupling between $|11\rangle$ and $|02\rangle$ states used for the CZ gate. (b) During a CZ gate, other neighboring qubits must be detuned to mitigate stray interactions. (c) Stray coupling between next-nearest neighbor qubits, approximately 1/20 of nearest neighbor coupling. A small detuning Δ is sufficient to avoid interaction. (d) Qubit idling frequencies to turn off nearest and next-nearest neighbor interactions. (e-g) Three step sequence to perform all 8 CZ of the repetition code with stray interactions in mind.

resource intensive in terms of frequency space. Instead, we use a three step sequence which naturally mitigates stray interactions. Figure E.24d shows the qubits in their idling state to minimize all interactions; neighboring qubits are detuned by $\Delta \approx 800$ MHz, next-nearest neighbor qubits are detuning by $\Delta \approx 100$ MHz. Additionally, this configuration is convenient to minimize microwave crosstalk between next-nearest neighbors, as resonant stray microwaves are detrimental to fidelity [**?**].

Figure E.24e-g shows the three step operation to perform all eight CZ gates, where similar Δs are maintained between non-interacting qubits. Additionally, this basic pattern is scalable to a one-dimensional array of arbitrary length. We note that these essential techniques can be transferred to the operation of the two-dimensional surface code where each qubit has four nearest neighbors, instead of two.

E.16.2 Evaluating the qubit spectrum

With generic operation principles outlined in the previous sections, we must choose specific operation frequencies for qubit idling, gates, and readout. In our Xmon qubits, the energy relaxation time T_1 varies over frequency; this is due to the fine structure from spectral distribution of incoherent defects [**?**]. Typically, T_1 times can differ by a factor of two to three over nearby frequencies, and careful characterization of the qubit spectrum is critical.

We see three kinds of features in the qubit spectrum. The most innocuous are incoherent defects, which suppress T_1 to the few μs range, but generally have small spectral width and can be safely swept past in frequency. We also see strongly and weakly coupled coherent

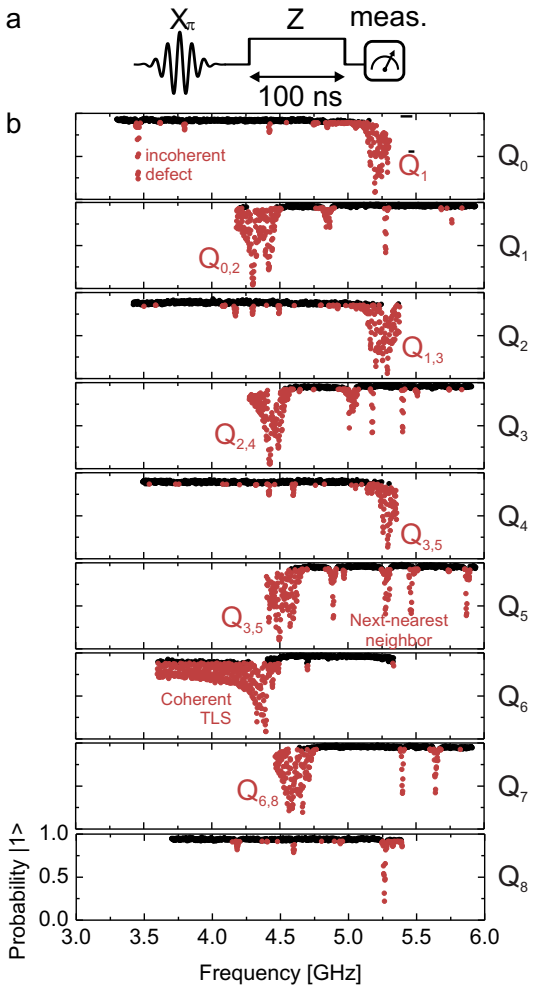


Figure E.25: **Qubit spectrum.** (a) Experiment for measuring the qubit spectrum. A qubit is excited to the $|1\rangle$ state and allowed to idle at different frequencies for 100 ns via a detuning pulse. (b) Spectrum data for nine qubits. Data is thresholded into good (black) and bad (red) regions. We define the threshold as 2% below the median $|1\rangle$ state population. This successfully identifies incoherent defects, coherent defects, and neighbouring as well as next-nearest neighbouring qubits.

features. Strongly coupled defects and strongly coupled qubits can coherently move population and affect phases – much more detrimental to algorithms – and sweeping qubits past them fast enough becomes challenging. Lastly, we see weakly coupled coherent defects and qubits which are slow enough such that we can move past them quickly in frequency. Thus, the qubits can be thought of as moving freely in a frequency band between strongly coherently coupled features, where operations should be done at frequencies away from incoherent defects and weakly coupled features.

Figure E.25a shows the experiment for measuring the qubit spectrum. The qubit is excited to the $|1\rangle$ state, idles at a frequency for 100 ns, and measured. By sweeping over the operable frequency range for a qubit, we can identify population loss from incoherent or coherent features. Figure E.25b shows the spectral data for the nine qubits. We identify bad regions in the spectrum by finding all frequencies that have population below a threshold value. We choose a threshold of 2% below the median $|1\rangle$ state population, corresponding to an added 5 μs T_1 mechanism over the qubit baseline. The data above threshold (operable regions) are plotted in black, and the data below threshold (inoperable regions) are plotted in red. Incoherent defects and weakly coupled coherent features (such as next-nearest neighbor qubits) have a thin spectral width and are easily avoided. Strongly coherently coupled features are easily identified through the coherent population swapping, such as seen in the spectrum of Q_6 at 4.3 GHz to 3.5 GHz and below.

E.16.3 Programming the repetition code

With the principles for operating gates and efficiently using the qubit spectrum above, we now construct the repetition code algorithm. Figure E.26a shows the high-level operation sequence for the nine qubit repetition code. Each measure qubit performs two CNOT gates with the control on the neighboring data qubits, and a measure. Figure E.26a shows the actual physical gate sequence. For our system we decompose CNOT gates into CZ and $\pi/2$ gates. Additionally, we use detune gates to move unused qubits away in frequency space to avoid unwanted interactions, such as in Fig. E.24. Echo X_π gates are inserted between CZ and detune gates to suppress non-Markovian noise [**?**]. Figure E.26b shows raw pulse waveforms for one repetition code cycle. The operations on each qubit have an XY, and Z control, as well as a multiplexed readout (RO) line. XY and RO waveforms are shown before up-conversion to GHz frequencies with an IQ mixer. Full control waveform data for eight cycles of the nine qubit repetition code can be found in Fig E.27.

Operating frequencies for idling, readout, and CZ gates are chosen away from frequencies characterized to have poor coherence, as in section E.16.2. Figure E.28 shows qubit frequencies at various stages in the gate sequence. The vertical bar associated with each qubit shows operable (grey) and non-operable (red) frequencies. Note that during a CZ, the qubit higher in frequency has its $|2\rangle$ state virtually populated, thus it is important to this state to also avoid non-operable regions. We also plot the AC stark shift $vs.$ time, as the qubit follows a non-trivial trajectory in frequency during readout, see Fig. E.28.

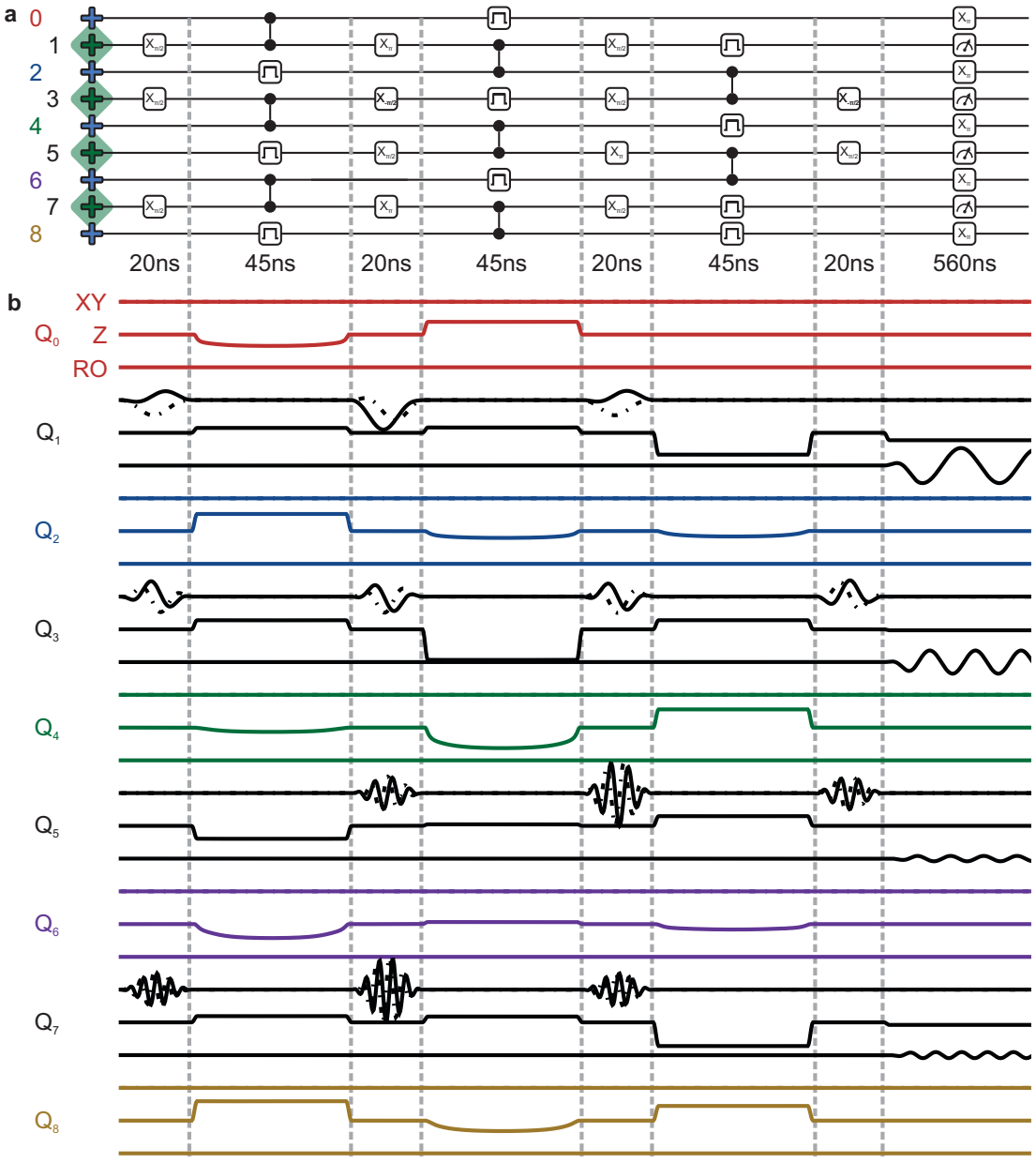


Figure E.26: **Single repetition code waveform data.** Rectangular gates indicate detuning pulses.

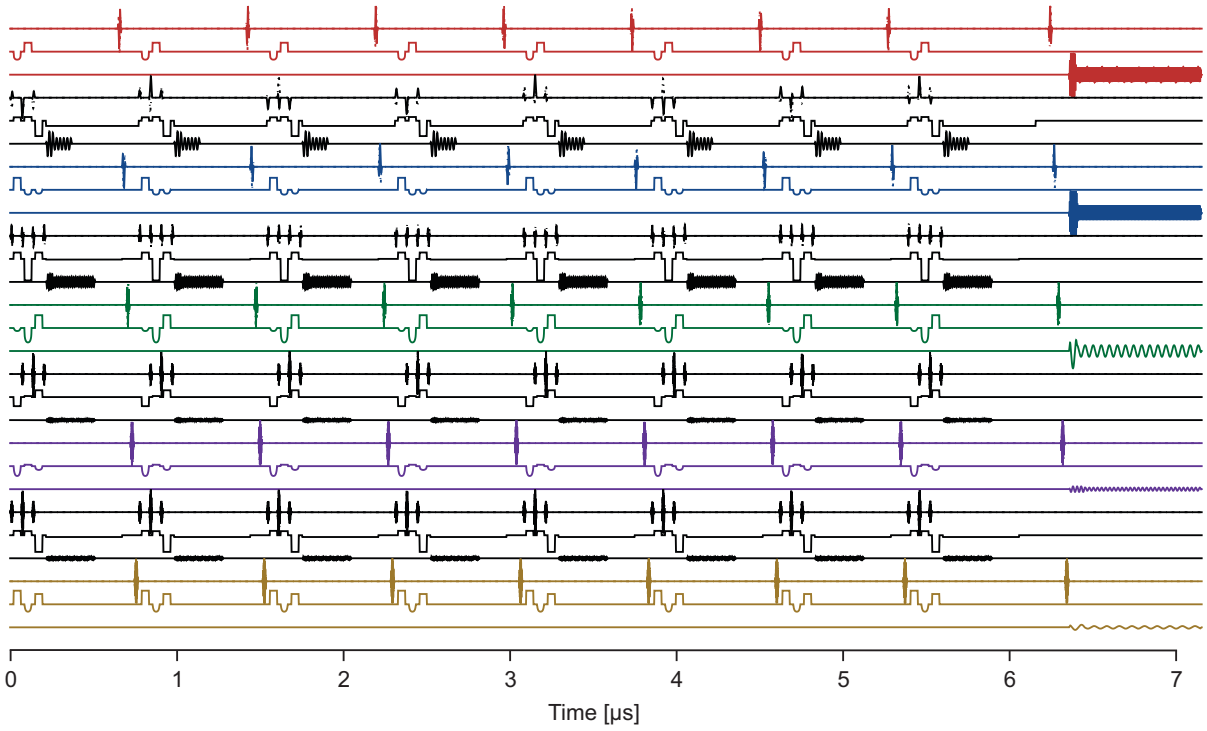


Figure E.27: **Waveform data for eight cycles of the nine qubit repetition code.**

E.17 Experimental setup

The experimental setup is described in Figure E.29.

E.18 Protecting the GHZ state from bit-flip errors: Conditional quantum state tomography

The density matrices of the GHZ states at the input and at the output of two cycles of the repetition code are reconstructed using quantum state tomography. We apply gates from { I, X/2, Y/2, X $\}^{\otimes 3}$. With quadratic maximum likelihood estimation, using the MATLAB packages SeDuMi and YALMIP, we constrain the density matrix to be physical. Non-idealities in

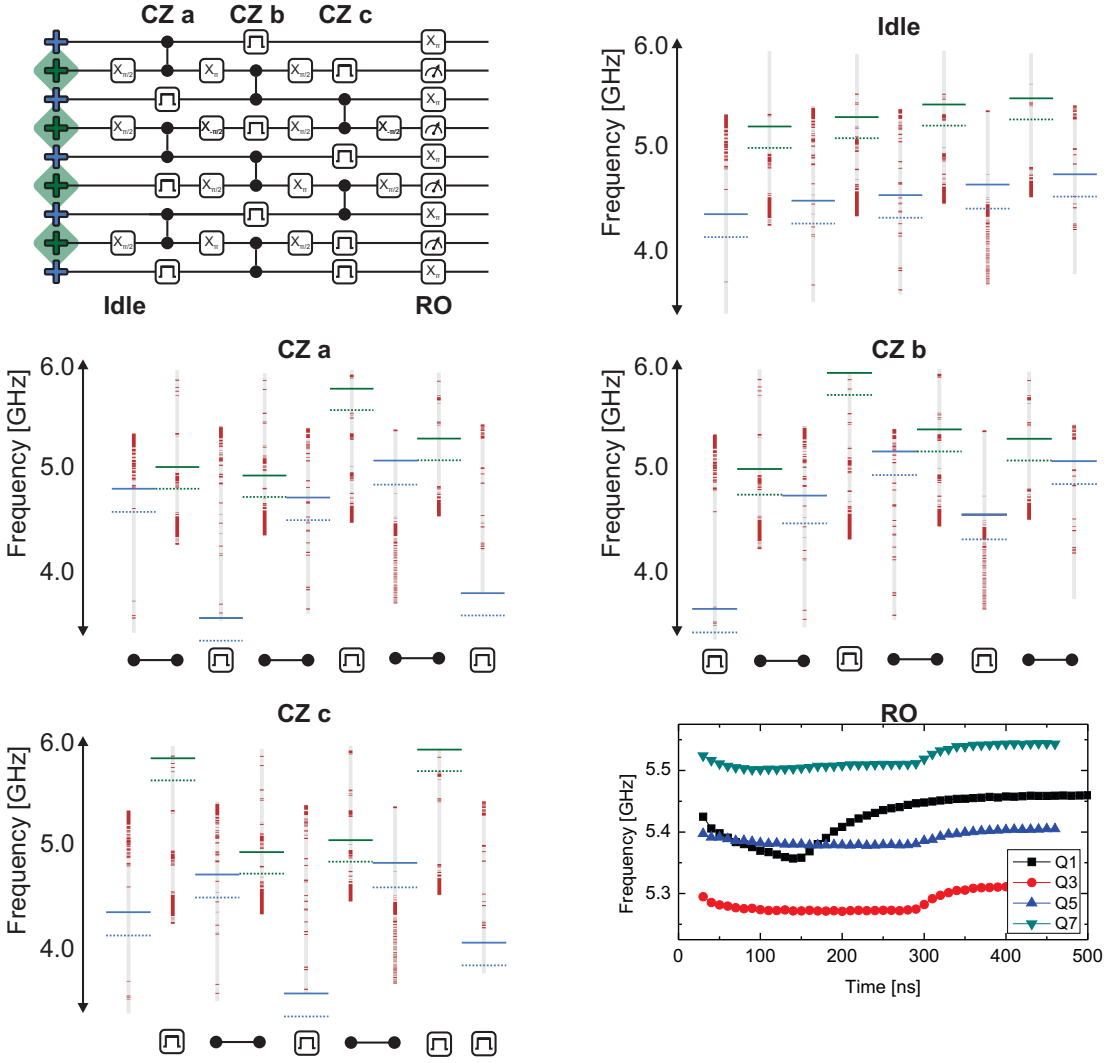
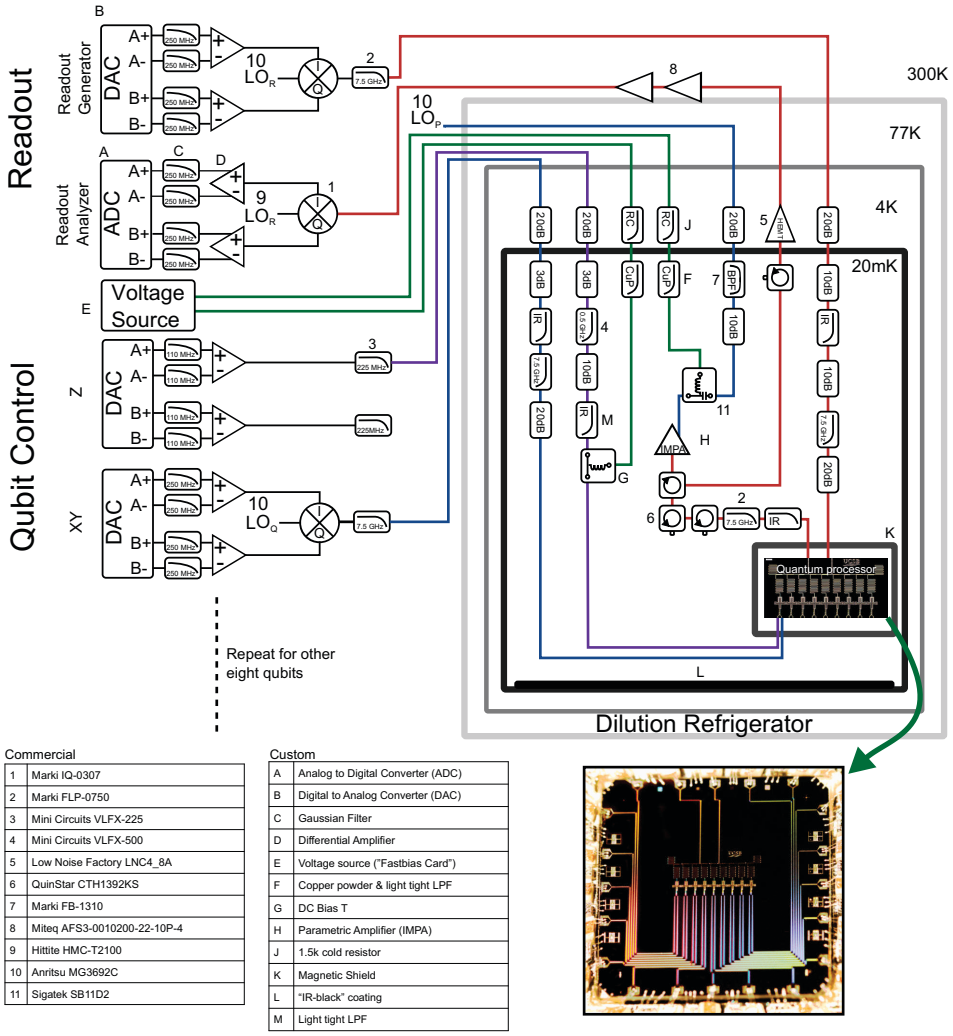


Figure E.28: **Qubit frequencies during the repetition code.** Vertical bars on qubit levels indicate operable regions (grey) and poor regions (red). Qubits are detuned to various frequencies during operation of the repetition code. Rectangular gates indicate detuning pulses. Different frequencies are chosen for idling, CZs and RO. Frequencies must be chosen to both reduce stray interactions (section E.16.1) and microwave crosstalk, while also avoiding frequencies with poor coherence, see section E.16.2.



Commercial

1	Marki IQ-0307
2	Marki FLP-0750
3	Mini Circuits VLFX-225
4	Mini Circuits VLFX-500
5	Low Noise Factory LNC4_8A
6	QuinStar CTH1392KS
7	Marki FB-1310
8	Miteq AFS3-0010200-22-10P-4
9	Hittite HMC-T2100
10	Anritsu MG3692C
11	Sigatek SB11D2

Custom

A	Analog to Digital Converter (ADC)
B	Digital to Analog Converter (DAC)
C	Gaussian Filter
D	Differential Amplifier
E	Voltage source ("Fastbias Card")
F	Copper powder & light tight LPF
G	DC Bias T
H	Parametric Amplifier (IMPA)
J	1.5k cold resistor
K	Magnetic Shield
L	"IR-black" coating
M	Light tight LPF

Figure E.29: **Electronics and Control Wiring.** Diagram detailing all of the control electronics, control wiring, and filtering for the experimental setup. Each qubit uses one digital to analog converter (DAC) channel for each of the X, Y, and Z rotations. Additionally, we use a DC bias tee to connect a voltage source to each qubit frequency control line to give a static frequency offset. All nine qubits are read out using frequency-domain multiplexing on a single measurement line. The readout DAC generates nine measurement tones at the distinct frequencies corresponding to each qubit's readout resonator. The signal is amplified by a wideband parametric amplifier [**?**], a high electron mobility transistor (HEMT), and room temperature amplifiers before demodulation and state discrimination by the analog to digital converter (ADC). All control wires go through various stages of attenuation and filtering to prevent unwanted signals from disturbing the quantum processor. Two local oscillators (LO_Q) are used for qubit XY control, at 4.38 and 5.202 GHz. The readout LO_R is at 6.58 GHz. All LO, DAC, and ADC electronics are locked to a 10 MHz SRS FS725 rubidium frequency standard. Photograph of nine qubit device shown in the lower right.

data qubit measurement and state preparation are suppressed by performing tomography on a zero-time idle [**?** **?**]. For conditional tomography, we separate out the measured data qubit probabilities based on the detection events. The experiment was repeated $12 \cdot 10^3$ times.

The raw and corrected output density matrices are shown in Fig. E.30.

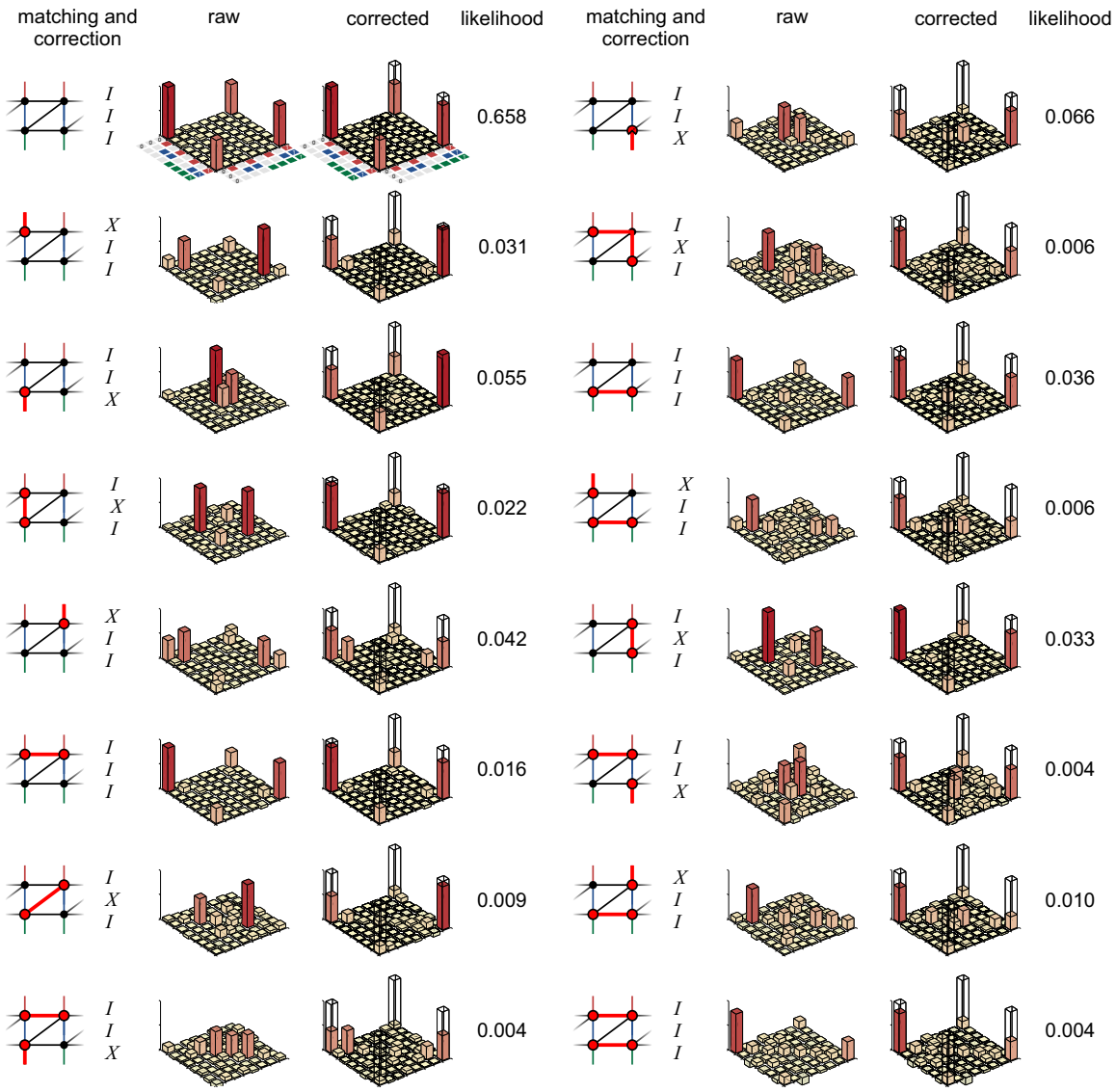


Figure E.30: **Protecting the GHZ state from bit-flip errors: All detection events.** The leftmost column displays the detection events and matching, indicating the most likely error, as well as the correction to apply in postprocessing to recover the input state. Raw and corrected output density matrices, reconstructed using quantum state tomography, conditional on all detection events. Corrected output density matrices are obtained by exchanging raw density matrix elements based on the correction. The real parts are shown. The likelihood indicates the prevalence of the detection event. See Fig. 4.5 for the quantum circuit diagram.

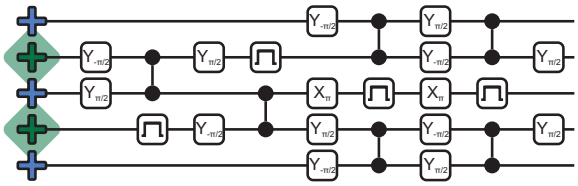


Figure E.31: **Quantum circuit for the generation of the GHZ state.** Rectangular gates indicate detuning pulses.

E.19 GHZ generation

The gate sequence for generation of the GHZ state used in the main text is shown in Fig. E.31.

Appendix F

Fabrication

F.1 Full cleanroom process

1. Control wiring

   (a) Grow MBE Al as in section F.3.

   (b) Optical lithography as in section F.5.

   (c) ICP etch as in table F.1.

2. Crossovers

   (a) SiO_2 deposition as in section F.4.

   (b) Al top wiring deposition as in section F.4.

3. Au alignment marks

(a) Deposit Ti/Au 10 nm / 150 nm stackup in liftoff process as alignment marks for e-beam lithography.

4. Xmon capacitor etch

   (a) Optical lithography as in section F.5.

   (b) ICP etch as in table F.1.

5. Dice wafer into quarters

6. Josephson junctions (one quarter at a time)

   (a) Spin e-beam reisst as in section F.2.1.

   (b) Expose design of section F.2.5 with exposure parameters from section F.2.2.

   (c) Deposit junctions as in section F.2.4.

This process was used to fabricate devices used in references [**? ? ?**].

F.2 Josephson junctions

We developed a Josephson junction process based off of the standard Dolan bridge technique, with enough structural integrity to build bridges up to 4 μm long and withstand an Ar-ion mill which removes native AlO_x from Al films used for making good contact with pre-patterned Al layers. Our process is designed to be used as the final step of device fabrication, where the Ar-ion mill is necessary to make good contact between the junction and the pre-patterned Al

that makes up the qubit capacitor. This is in contrast to other groups who deposit the junction and Al that makes up the qubit capacitor in the same step.

F.2.1 Resist stackup

We use a 500nm MAA EL9 and 300nm PMMA 950K A4 resist bilayer. Thicknesses were verified using a Woolam M2000DI variable angle spectroscopic ellipsometer. The PMMA thickness of 300nm is more than what is generally used in the field. We do this to improve the structural integrity of the Dolan bridge. The recipe is as follows:

1. Spin MAA at 1500 rpm for 45 seconds.
2. Bake at 160° C for 10 minutes.
3. Spin PMMA at 2000 rpm for 40 seconds.
4. Bake at 160° C for 10 minutes.
5. Deposit 10nm of Au at 0.05 nm/s in NRC 3117 three source thermal evaporator. Au acts as a conduction layer to prevent charging of sapphire substrate during e-beam lithography.
6. Expose in e-beam writer, see section F.2.2.
7. Strip Au in Gold etchant type TFA (KI 18-42 ppm solution) for 10 seconds, DI rinse.
8. Develop in 3:1 IPA,MIBK mixture for 45 seconds with gentle agitation.
9. Immerse in IPA for 10 seconds with gentle agitation.

10. Blow drop with N_2 at 10 PSI until dry.

F.2.2 E-beam writer

For e-beam writing, we use a 100kV JEOL JBX-6300FS system. This system, combined with our resist stackup allows us to either expose only the MAA layer, or to expose both the PMMA and MAA. By doing this, we can directly define the undercut for the Dolan bridge and other features. We use the following doses:

- MAA dose - 350 μC/cm^2
- PMMA+MAA dose - 1500 μC/cm^2
- Clearing dose - 2000 μC/cm^2

Where the strong clearing dose is used to ensure resist is cleared from contact pads. See figure F.1 for a SEM image of the developed resist profile.

F.2.3 Argon ion mill

In order to make contact with a pre-patterned Al layer, we use an Ar-ion mill to remove the native oxide. We use a Kaufman source with beam energy of 400 eV, 21 mA current, and width of 3.2" at the substrate for a dose of $\sim$ 2.6 x 10^{15}cm^{-2}s^{-1}.

To calibrate the Ar-ion mill etch rate, we take a set of pre-fabricated dies made for resonator experiments. These devices have feedlines consisting of 100 thick nm of Al that span the chip. We measure how much material the mill removes from a die by measuring the resistance of

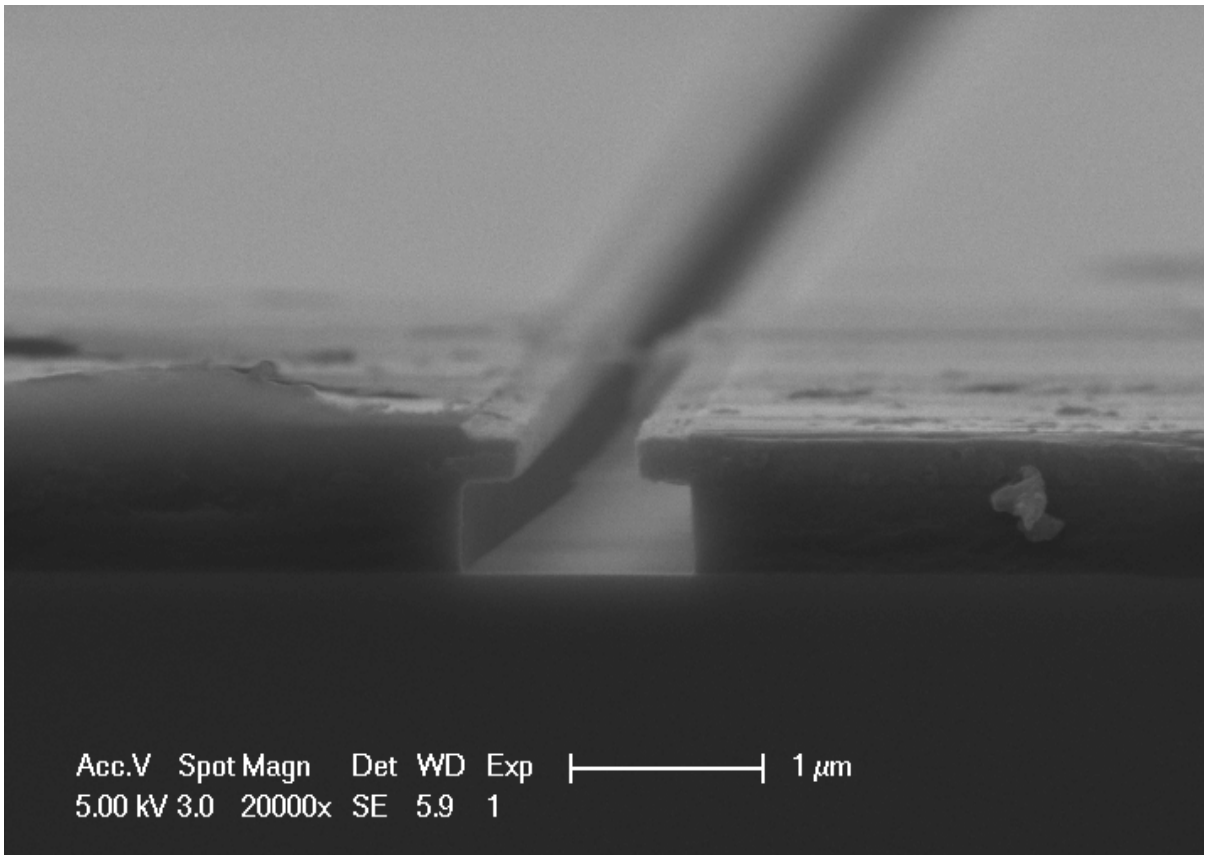


Figure F.1: **E-beam resist stackup: side profile.** SEM image of e-beam resist stackup. We see a clear undercut of 100-200nm as defined in the write file. We find the MAA sidewalls to be very vertical, indicating little pattern bleeding. We find thinning in the PMMA layer after development, but not enough to cause failure. No visible MAA adheres to the substrate.

the feedline before and after milling for a set time. This method naturally averages over a large area. Given that Al and AlO_x have different etch rates, testing a variety of different mill times we can determine when all oxide is removed, see Fig. F.2. We chose a mill time of 3 minutes 30 seconds, comfortably over the time it takes to remove all oxide. This ensures that any minor system drift will not prevent good contact. Critical current measurements were done in Ref. [**?**].

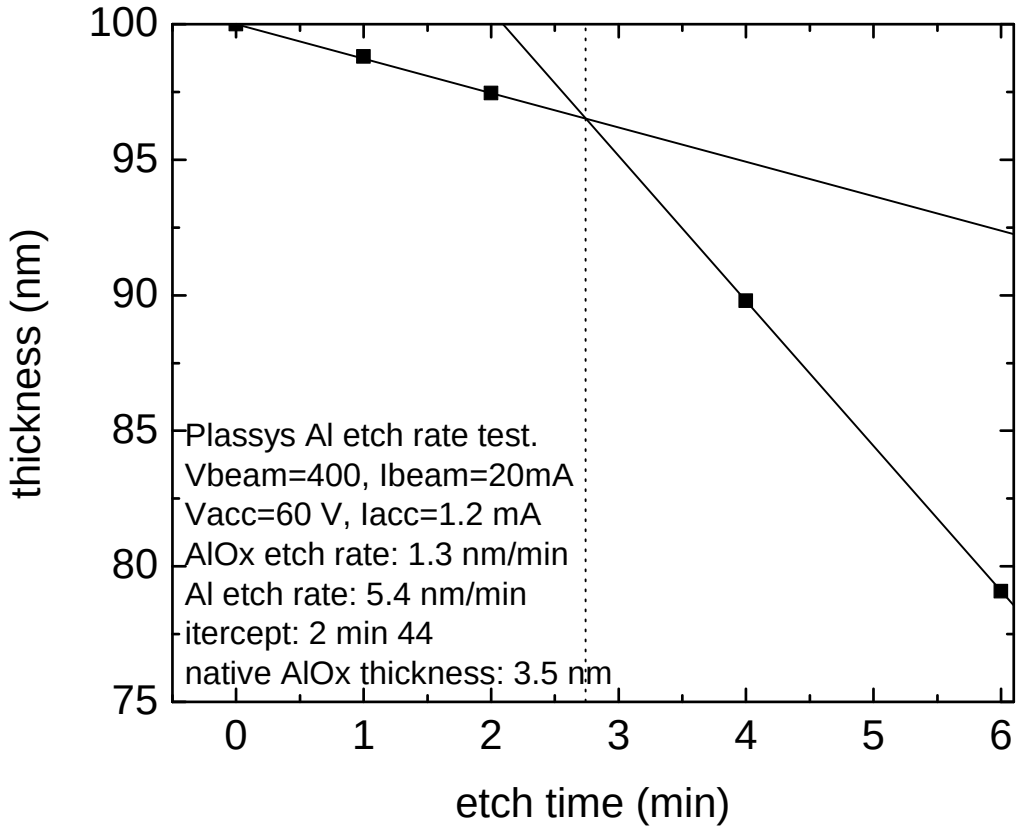


Figure F.2: **Ar-ion mill calibration.** Measured thickness of an initial 100nm thick film vs etch time. Change in thickness determined through resistance measurements. Al and AlO_x mill rates are fit from the data.

F.2.4 Deposition process

The deposition is a double angle process done in a Plassys vacuum system.

1. Pump down overnight to let resist outgas. Generally 17 hours with final pressure $< 10^{-7}$ mbar.
2. Ar-ion mill for 3 minutes 30 seconds to remove native oxide.
3. Ti getter for 3 minutes at 0.2nm/s to clean chamber.
4. Deposit 65 nm of Al at 1 nm/s at 62.2° from normal to the substrate (30 nm film on substrate).
5. Oxidize at 5 mbar for 30 minutes.
6. Deposit 100 nm of Al at 1 nm/s normal to the substrate.
7. Oxidize at 5 mbar for 5 minutes.
8. Remove from Plassys.
9. Soak in Shipley 1165 (N-methyl-2-pyrrolidine) at 80° for 3 hours.
10. Clean with sonication, IPA, spin dry.

F.2.5 Junction design

We use a two perpendicular lines which cross to form the Josephson junction. This makes the junction area first order insensitive to alignment error. We add a stress relief "T" shape

to one of the junction leads to improve Dolan bridge reliability. Without the T, we find that smaller junctions often fail. We hypothesize that the T elongates the bridge and removes a stress concentrator. We keep the length of the bridge fixed when we change the junction width for junctions with width $< 1\mu$m. This way, we can optimize a single bridge and use it for most junction sizes. Junction height can range from 0.1-0.5μm, junction width can range from 0.1-4μm. This allows junction areas to range by two orders of magnitude. We typically find an increase of 50 nm in the junction height as compared to the design width, and 70 nm in the junction height, meaning the physical junction is larger than designed.

F.2.6 Reliability

We find this process to be extremely reliable. 76 transmon qubits made with our junction process between 2011-2014 have had 100% yield. Other members of the group have used this process to yield devices with thousands of junctions. We also find that our junctions do not age over the time span of days, as other groups have seen. This is likely due to the ion mill removing organic residue before deposition, and the long liftoff in 80° solvent which anneals the junction if it were to anneal. Resistance data showing this can be found in Fig.F.5. The 9 qubit device used in Chapter 8 has four and five qubits with maximum frequencies around 5.9 and 5.3 GHz respectively. The relative standard deviation is 0.3% and 0.7% respectively in qubit frequency, with a larger variance for smaller junction area. This indicates junction areas are controlled to 5% and 8% respectively, as the qubit frequency depends on the square root of the junction area.

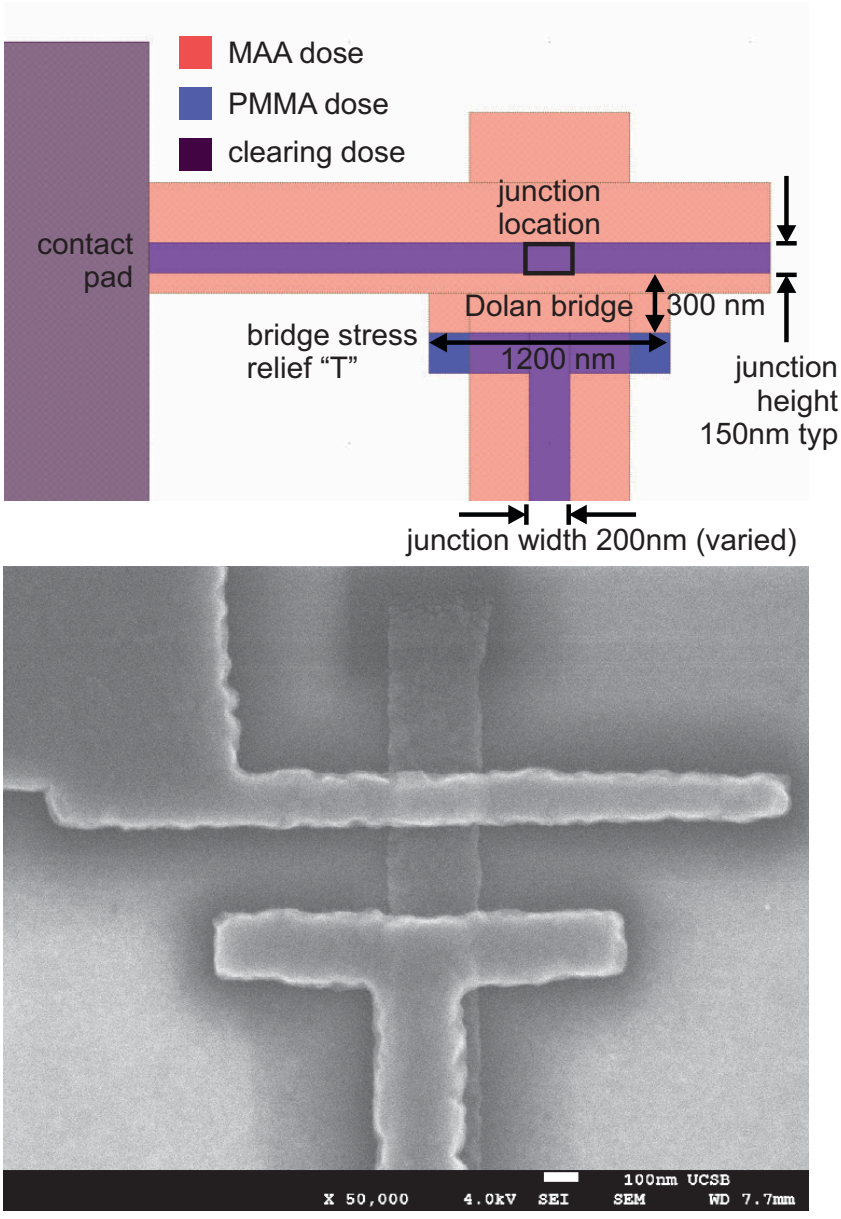


Figure F.3: **Junction design and SEM image.**

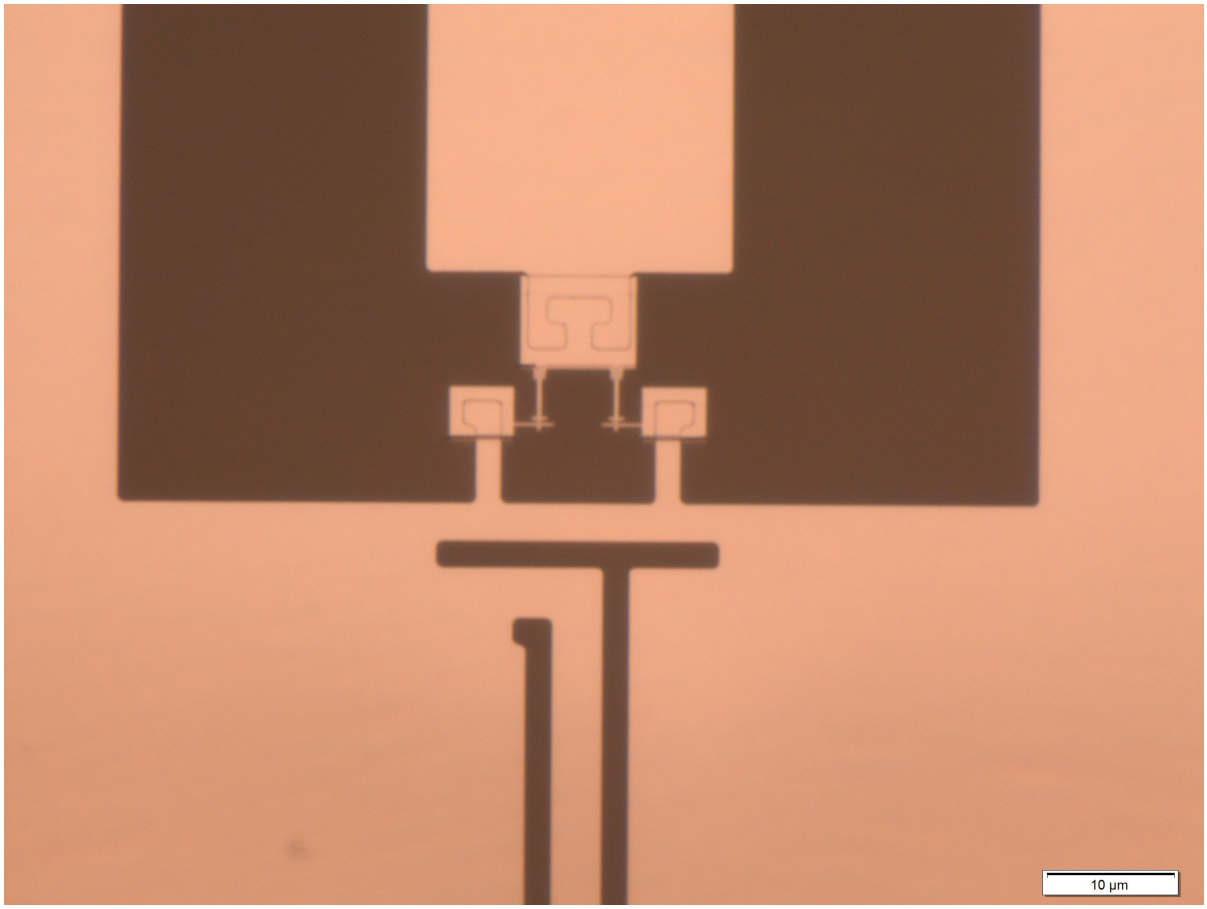


Figure F.4: **Optical image of the Xmon SQUID loop with deposited junctions.**

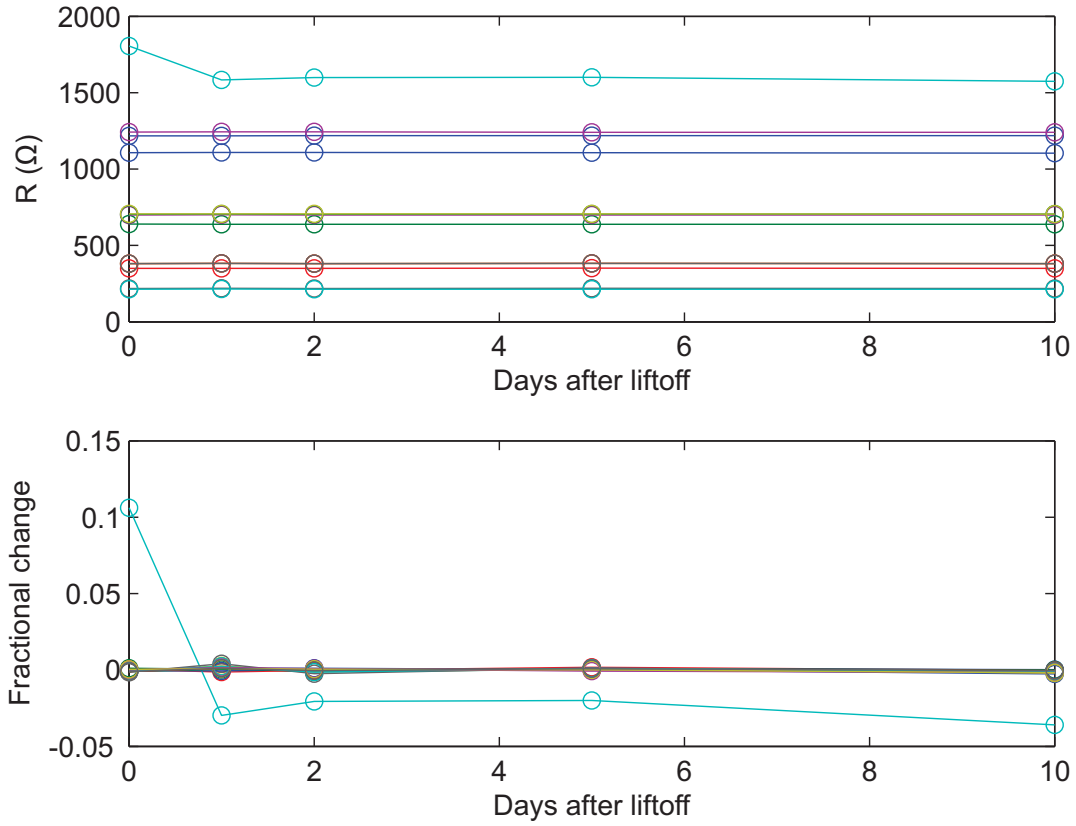


Figure F.5: **Junction aging.** Test junctions resistances are measured vs. time after fabrication. Minimal aging is observed.

Table F.1: BCl_3Cl_2 etch recipe for Al films.

	Step 1	Step 2	Step 3	Step 4	Step 5
BCl_3	20 SCCM	20 SCCM	20 SCCM	0 SCCM	0 SCCM
Cl_2	40 SCCM	40 SCCM	40 SCCM	0 SCCM	0 SCCM
CF_4	0 SCCM	0 SCCM	0 SCCM	50 SCCM	50 SCCM
Pre.	3.0 Pa	0.7 Pa	0.7 Pa	2.0 Pa	2.0 Pa
SRC FWD	300 W	300 W	300 W	700 W	700 W
BIAS FWD	0 W	0 W	70 W	0 W	20 W
Step Time	5 s	5 s	33 s	5 s	5 s

F.3 High quality capacitors

We used molecular beam epitaxy (MBE) Al deposited on a *c*-plane sapphire substrate. The Al film thickness is approximately 100 nm, deposited at room temperature. The sapphire substrate was cleaned by load-lock outgassing at $200\,^{\circ}\mathrm{C}$, followed by heating to $850\,^{\circ}\mathrm{C}$ in $\sim 10^{-6}$ Torr activated oxygen, identical to the process outlined in Ref. [**?**].

Etching of the Xmon capacitor is done with a Panasonic E626I inductively coupled plasma (ICP) dry etcher, using a BCl_3Cl_2 etch, shown in table F.1. Immediately after etching, wafer is immersed in DI water for 10 minutes to passivate any residual Cl. Photoresist is cleaned via sonication in Acetone, IPA and spin dry.

F.4 Crossovers

We add crossovers to our devices to improve ground plane connectivity and remove slotline modes. We use a SiO_2 dielectric which has a significant loss tangent. To prevent added decoherence to our qubits, we design our Xmon qubits with intrinsic ground plane connectivity and add crossovers to the 50 Ω wiring used for control. We use a two-step liftoff process to add

crossovers.

SiO_2:

1. Begin with wafer that has CPW control lines etched.
2. Expose and develop SPR 955 0.9 μm resist.
3. Deposit 200 nm SiO_2 at 0.3 nm/s in custom Temescal e-beam evaporation system.
4. Liftoff in 80° Shipley 1165 (NMP).
5. Clean with sonication, IPA, spin dry.

Al:

1. Expose and develop SPR 955 0.9 μm resist.
2. Ar-ion mill in Plassys system to remove native AlO_x.
3. Deposit 200 nm Al in Plassys system.
4. Liftoff in 80° Shipley 1165 (NMP).
5. Clean with sonication, IPA, spin dry.

F.5 Optical lithography

1. Cover wafer with HMDS for 30 s while on spinner.
2. Spin at 3000 RPM for 45 s.

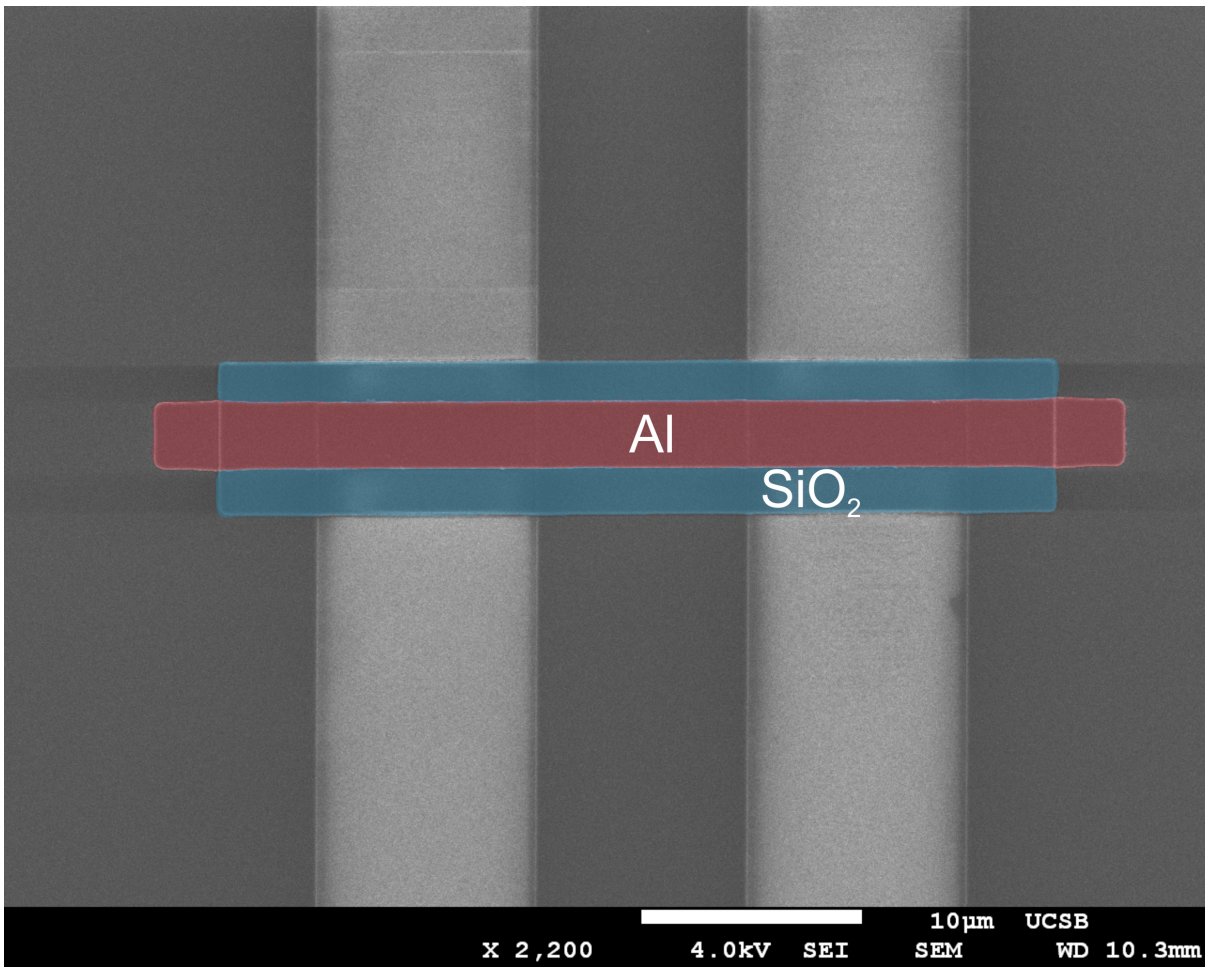


Figure F.6: **False color SEM image of crossover.** Al (red) on SiO_2(blue) crossover, over sapphire substrate (light grey), bridging groundplanes (dark grey).

3. Change spinners as HMDS vapor can contaminate resist.

4. Spin SPR 955 0.9 μm resist at 3000 RPM for 45 s.

5. Pre-exposure bake for 90° C for 90 s.

6. Expose in GCA 200 I-Line Wafer Stepper for 0.37 s.

7. Post-exposure bake at 110° C for 60 s.

8. Develop in AZ 300 MIF developer with gentle agitation for 60 s.

9. Immerse in DI water with gentle agitation for 60 s.

10. N_2 blow dry at 20 psi until dry.

SPR 955 0.9 μm photoresist is filtered with 0.45 μm filter. This is critical for yield on complex, small featured size devices.